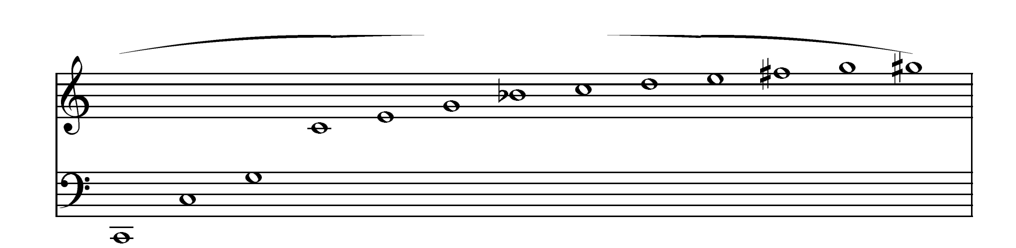


Module d'ouverture
Deuxième Année



SON ET MUSIQUE

Philippe GUILLAUME

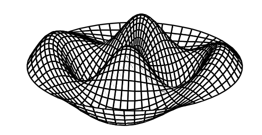


Table des matières

Introduction

Ce cours s'adresse aux étudiants de deuxième année de l'INSA Toulouse. Il a été créé à l'occasion d'une réforme pédagogique mise en place en septembre 2002, qui a introduit en particulier des « modules d'ouverture » censés apporter de la transversalité aussi bien que développer la culture générale. C'est dans cet esprit que ce cours a été conçu. La nature de l'objet étudié fait que l'on aborde tout naturellement des notions de physique (génération et propagation des sons), de mathématiques (analyse de Fourier), de psychoacoustique (étude de la perception des sons, qui se différencie de ce que peut donner une mesure physique), de théorie du signal (analogique et numérique), d'algorithmique et d'informatique (compression MP3 des sons), et bien entendu... de musique !

L'objectif de ce cours est de donner un aperçu global sur le son musical, depuis sa production par les instruments de musique classiques jusqu'aux sons obtenus par synthèse numérique, sans chercher pour autant l'exhaustivité : il ne s'agit pas d'un catalogue, mais, je l'espère, d'une porte d'entrée permettant d'aborder ensuite sans difficulté tout autre thème du même domaine. Après une initiation à l'acoustique et une étude de différents types de sons au chapitre 1, nous aborderons au chapitre 2 les modes vibratoires de quelques instruments, l'étude de leur timbre, et ferons le lien avec la question des gammes (tempéraments) au chapitre 3. À la suite d'une excursion psychoacoustique au chapitre 4, qui sera l'occasion de faire la connaissance d'une belle illusion acoustique, nous traiterons plusieurs aspects du son numérique dans les chapitres 5 et 6 : l'échantillonnage, la compression basée sur les propriétés de l'audition (MP3), les effets sonores (simulation de la réverbération d'une salle, effet Leslie) et les sons de synthèse comme par exemple la technique FM de Chowning, rendue très populaire par la série des DX7.

Le mode d'emploi de ce cours est : papier-crayon + ordinateur + instrument de musique. *Papier-crayon*, car il s'agit d'un cours scientifique, et pour bien comprendre et assimiler les notions abordées, il est nécessaire de refaire certains calculs par soi-même, en changeant les données ou les hypothèses et en interprétant les variations obtenues dans les résultats. *Ordinateur*, car une bonne partie du cours concerne le son numérique, et l'ordinateur fournit un instrument formidable pour analyser des sons, faire des expériences avec eux (effets sonores) et créer de nouvelles sonorités. Et finalement, *instrument de musique*, car... vous devinerez par vous-même !

Remerciements. *Je remercie sincèrement Brigitte Bidegaray pour la lecture très attentive qu'elle a faite de ce document, ainsi que pour ses suggestions et anecdotes que j'ai insérées dans le texte sans plus de formalités ! Je tiens également à remercier chaleureusement tous les collègues de l'INSA qui m'ont encouragé dans la création de ce cours.*

Chapitre 1

Les sons

Le son et l'air sont intimement liés : il est bien connu que les luniens (les habitants de la lune) n'ont pas d'oreilles ! Aussi commencerons-nous l'étude du son par celle de son moyen de transport : l'air. Les sons qui se propagent dans notre atmosphère consistent en une variation de la pression $p(x, y, z, t)$ en fonction de la position x, y, z et du temps t. Ce sont ces variations de pression que notre oreille perçoit. Dans ce chapitre, nous étudierons d'abord la manière dont elles se propagent sous forme d'ondes. Nous décrirons ensuite quelques types de sons et plusieurs façons de les représenter. Nous terminerons par la notion de filtrage, qui permet par exemple d'isoler certaines fréquences.

1.1 Propagation des sons

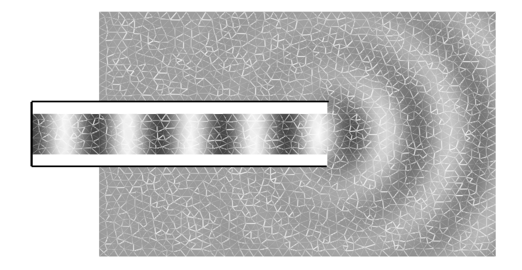


FIG. 1.1 – ondes de pression dans un tube ouvert à droite, pression imposée à gauche.

La propagation des ondes sonores se fait *a priori* dans toutes les directions, et dépend des obstacles rencontrés. Nous étudions essentiellement les *ondes planes*, c'est-à-dire ne dépendant que d'une seule direction d'espace. Nous supposons que cette direction est l'axe des x, et par conséquent la pression $p(x, y, z, t)$ ne dépend ni de y, ni de z. On peut donc la noter simplement $p(x, t)$. Cette forme de fonction représente une onde plane se propageant dans l'espace, mais également l'onde sonore à l'intérieur d'un tube (cf. fig. 1.1), comme par exemple celle qui se propage dans un tuyau d'orgue.

1.1.1 Un peu de modélisation physique

L'équation qui régit le déplacement des ondes sonores s'obtient à partir de trois équations fondamentales issues de la physique des milieux continus.

Conservation de la masse

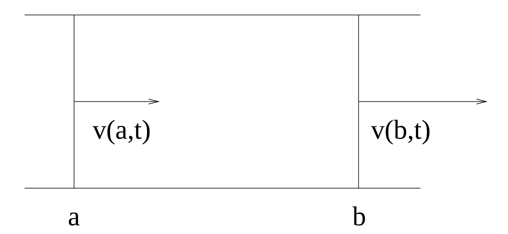


FIG. 1.2 – bilan de la masse dans la tranche d'air : il n'y a ni création ni disparition !

On note $\rho(x,t)$ la densité de l'air, et $v(x,t)$ la vitesse (moyenne) des molécules d'air qui se déplacent sous l'effet de l'onde sonore, à ne pas confondre avec la vitesse de propagation du son notée c. En l'absence de vent, ce que nous supposons ici, cette vitesse v oscille autour de 0, et p et ρ oscillent autour de leurs valeurs moyennes p_0 et ρ_0. Dans la tranche d'air de section S comprise entre a et b (cf. fig. 1.2), contenant la masse d'air $m(t)$, on égale le flux de masse $\Phi(t)$ qui rentre dans cette tranche et la variation (la dérivée) de cette masse :

$$\Phi(t) = m'(t),$$

avec

$$\begin{aligned} m(t) &= S \int_a^b \rho(x,t)\,dx, \\ \Phi(t) &= S[\rho(a,t)v(a,t) - \rho(b,t)v(b,t)], \end{aligned}$$

ce qui donne, en notant ∂_t la dérivée partielle par rapport au temps :

$$S[\rho(a,t)v(a,t) - \rho(b,t)v(b,t)] = S \int_a^b \partial_t \rho(x,t)\,dx.$$

En divisant par $b-a$ et en faisant tendre $b-a$ vers 0, on obtient après simplification par S

$$-\partial_x(\rho(x,t)v(x,t)) = \partial_t \rho(x,t). \tag{1.1}$$

L'hypothèse de l'acoustique linéaire consiste à supposer que les variations par rapport à l'état d'équilibre sont faibles (d'où le paramètre ε, supposé « petit») :

$$\begin{aligned} v(x,t) &= \varepsilon v_1(x,t), \\ \rho(x,t) &= \rho_0 + \varepsilon \rho_1(x,t). \end{aligned}$$

En substituant ces expressions dans (1.1) et en négligeant les termes en ε^2, on obtient alors l'équation de la conservation de la masse, appelée aussi *équation de continuité* :

$$\partial_t \rho_1(x,t) + \rho_0 \partial_x v_1(x,t) = 0. \tag{1.2}$$

Equation fondamentale de la dynamique

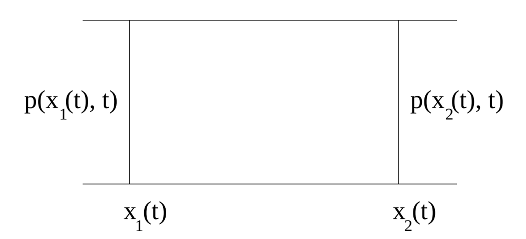


FIG. 1.3 – tranche d'air en pleine migration. Son accélération est la résultante des forces de pression.

Pour la tranche d'air comprise entre les abscisses $x_1 = x_1(t)$ et $x_2 = x_2(t)$ (cf. fig. 1.3) qui suivent le mouvement des molécules d'air, et vérifient donc $x_1'(t) = v(x_1(t), t)$ et $x_2'(t) = v(x_2(t), t)$, on écrit la relation fondamentale de la dynamique $F = d(mv)/dt$:

$$\begin{aligned}
S[p(x_1,t) - p(x_2,t)] &= \frac{d}{dt}\left(S\int_{x_1(t)}^{x_2(t)} \rho(x,t)v(x,t)dx\right) \\
&= S\Big(\rho(x_2,t)v(x_2,t)x_2'(t) - \rho(x_1,t)v(x_1,t)x_1'(t) \\
&\quad + \int_{x_1(t)}^{x_2(t)} \partial_t(\rho(x,t)v(x,t))dx\Big) \\
&= S\left(\rho(x_2,t)v^2(x_2,t) - \rho(x_1,t)v^2(x_1,t) + \int_{x_1(t)}^{x_2(t)} \partial_t(\rho(x,t)v(x,t))dx\right),
\end{aligned}$$

ce qui donne en divisant par $x_2 - x_1$ et en faisant tendre $x_2 - x_1$ vers 0 :

$$-\partial_x p(x,t) = \partial_x(\rho(x,t)v^2(x,t)) + \partial_t(\rho(x,t)v(x,t)).$$

De nouveau, si les variations par rapport à l'état d'équilibre sont petites, avec

$$p(x,t) = p_0 + \varepsilon p_1(x,t),$$

on obtient en négligeant les termes en ε^2 *l'équation d'Euler* :

$$-\partial_x p_1(x,t) = \rho_0 \partial_t v_1(x,t). \tag{1.3}$$

Equation d'état

En supposant qu'il n'y ait pas d'échange de chaleur entre tranches d'air ou avec l'extérieur (compression et détente adiabatiques car très rapides), *l'équation d'état* exprime que la variation de pression est proportionnelle à la variation de densité :

$$p_1(x,t) = c^2 \rho_1(x,t), \tag{1.4}$$

relation qui exprime aussi que l'air a un comportement élastique : il réagit comme un ressort. En substituant cette expression dans (1.2), on obtient une autre expression de l'équation d'état :

$$\partial_t p_1(x,t) + c^2 \rho_0 \partial_x v_1(x,t) = 0. \tag{1.5}$$

1.1.2 Equation des ondes

En dérivant par rapport au temps l'équation d'état (1.5) et par rapport à x l'équation d'Euler (1.3) on obtient

$$\partial_{t^2} p_1(x,t) = -c^2 \rho_0 \partial_{tx} v_1(x,t),$$
$$\partial_{x^2} p_1(x,t) = -\rho_0 \partial_{tx} v_1(x,t),$$

et en comparant ces deux relations, on en déduit *l'équation des ondes* :

$$\partial_{t^2} p_1(x,t) = c^2 \partial_{x^2} p_1(x,t). \tag{1.6}$$

On peut alors montrer que la solution générale de cette équation est de la forme

$$p_1(x,t) = g(x - ct) + h(x + ct).$$

La fonction $g(x - ct)$ est constante sur les points qui vérifient $x - ct =$ constante (une telle courbe $x(t)$ est appelée une *caractéristique*), et représente une onde *progressive* se propageant sur l'axe des x à la *vitesse du son* c de la gauche vers la droite. De même, la fonction $h(x+ct)$ est constante sur les points d'abscisse $x + ct =$ constante, et représente une onde progressive se propageant à la vitesse c de la droite vers la gauche. Pour l'air à la température T exprimée en degrés Kelvin (avec 0°C = 273K), les valeurs approximatives de vitesse du son, de densité et de pression atmosphérique (en *pascals* et en *bars*) sont

$$c = 20\sqrt{T}, \quad \rho_0 = \frac{353}{T}, \quad p_0 = 1.013\,10^5 \text{ Pa} = 1.013 \text{ bar à } 0°\text{C},$$
$$c = 330 \text{ m/s à } 0°\text{C}, \; c = 340 \text{ m/s à } 16°\text{C}.$$

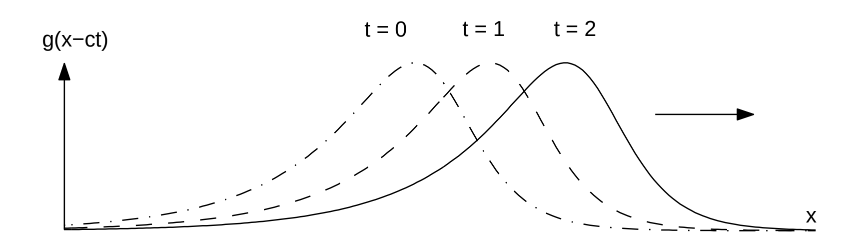


FIG. 1.4 – trois « photographies» d'une onde plane progressive le long d'un axe.

Par exemple, les fonctions

$$u_+(x,t) = \sin(kx - 2\pi f t),$$
$$u_-(x,t) = \sin(kx + 2\pi f t),$$

avec $k = 2\pi f/c$, sont des solutions de l'équation des ondes. Elles sont périodiques en temps et en espace, la période en espace

$$\lambda = \frac{2\pi}{k} = \frac{c}{f}$$

étant appelée la *longueur d'onde*. C'est une des formes les plus élémentaires de son musical de hauteur ou *fréquence* f, dont le timbre est proche de celui d'une flûte douce.

Ces deux fonctions u_+ et u_- progressent en sens inverse. En les additionnant, on obtient une nouvelle fonction intéressante, également solution de l'équation des ondes :

$$\begin{aligned} p_1(x,t) &= \sin(kx - 2\pi f t) + \sin(kx + 2\pi f t) \\ &= 2\sin(kx)\cos(2\pi f t). \end{aligned}$$

On constate qu'aux points $x = n\pi/k$, $n \in \mathbb{Z}$ (ensemble des entiers relatifs), la pression est constante (égale à p_0) : ces points sont appelés des *nœuds* de vibration, alors qu'aux points $x = (n + 1/2)\pi/k$, $n \in \mathbb{Z}$, la pression $p(x,t) = p_0 \pm 2\varepsilon\cos(2\pi f t)$ subit ses variations d'amplitude maximale : ces points sont appelés des *ventres* de vibration. De telles ondes sont dites stationnaires.

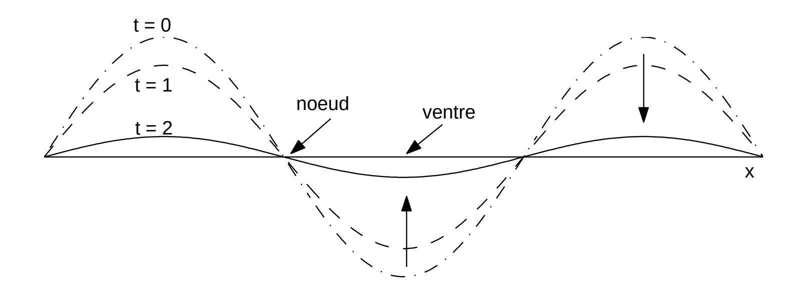


FIG. 1.5 – trois « photographies» d'une onde plane stationnaire.

1.1.3 Equation de Helmholtz

En physique, une onde ne comportant qu'une seule fréquence, c'est-à-dire de la forme

$$p_1(x,t) = \varphi(x)\exp(2i\pi f t)$$

où φ peut aussi être une fonction complexe[1] et $f \in \mathbb{R}$ (ensemble des nombres réels) est dite *harmonique*. Les parties réelles ou imaginaires d'une telle onde sont également harmoniques.

[1]L'usage des nombres et fonctions complexes est à la fois pratique et courant. Pour obtenir le signal physique associé à une fonction complexe, il suffit d'en prendre la partie réelle. Le signe de f indique le sens de parcours du cercle trigonométrique. Lorsque l'on passe dans le domaine réel, sachant que $\cos(-x) = \cos(x)$ et $\sin(-x) = -\sin(x)$, on peut toujours supposer que la fréquence f est ≥ 0, quitte à changer le signe du sinus.

D'une manière générale, on appelle fonctions *à variables séparées* les fonctions de la forme

$$p_1(x,t) = \varphi(x)\psi(t). \tag{1.7}$$

Si de plus φ est réelle, on dit que l'onde est *stationnaire* : à un coefficient réel multiplicatif près $\varphi(x)$, tous les points subissent simultanément la même variation de pression $\psi(t)$. Nous retrouverons ces ondes dans l'étude des cordes, des lames et des tubes.

En reportant l'équation (1.7) dans (1.6), on obtient après division par $\varphi(x)\psi(t)$

$$\frac{\psi''(t)}{\psi(t)} = c^2 \frac{\varphi''(x)}{\varphi(x)},$$

expression qui visiblement ne peut être qu'une constante. Si on pose que cette constante vaut $-(2\pi f)^2$ avec f réel arbitraire[2], on obtient d'une part

$$\psi''(t) = -(2\pi f)^2 \psi(t),$$

dont la solution générale est $\psi(t) = A\exp(2i\pi ft) + B\exp(-2i\pi ft)$. Si $B = 0$ ou $A = 0$, l'onde est harmonique de fréquence $\pm f$. D'autre part, pour $k = 2\pi f/c$, on obtient *l'équation de Helmholtz* homogène (c'est-à-dire sans autre terme que « l'inconnue» φ) :

$$\varphi''(x) + k^2\varphi(x) = 0, \tag{1.8}$$

dont la solution générale est $\varphi(x) = \alpha\exp(ikx) + \beta\exp(-ikx)$.

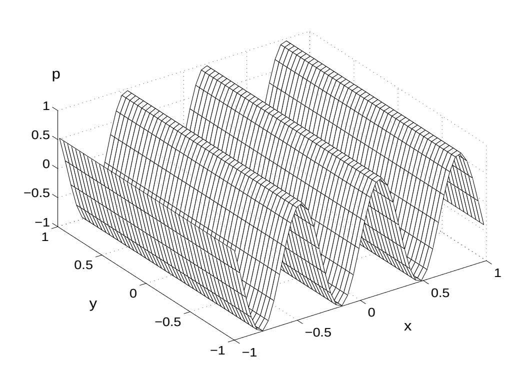


FIG. 1.6 – une onde plane harmonique. Elle se propage le long de l'axe Ox (coupe 2D) sans amortissement.

Ainsi, les ondes de pression harmoniques de fréquence f sont de la forme

$$p_1(x,t) = [\alpha\exp(ikx) + \beta\exp(-ikx)]\exp(2i\pi ft)$$

où α, $\beta \in \mathbb{C}$ (ensemble des nombres complexes) seront déterminés par les conditions aux limites (cf. chapitre 2), tandis que les ondes harmoniques stationnaires de fréquence f sont de la forme

$$p_1(x,t) = \alpha\sin(k(x-x_0))\exp(2i\pi ft).$$

[2]Le choix de la constante *négative* nous permet de sélectionner les solutions physiques qui nous intéressent : celles qui sont sinusoïdales en temps. Il existe d'autres solutions, mais nous ne les utiliserons pas.

En appliquant la même démarche (*i.e.*, en partant des trois équations fondamentales), on montre que dans le cas général, où les ondes ne sont plus forcément des ondes planes, la pression est solution de l'équation des ondes à trois dimensions d'espace

$$\partial_{t^2} p_1(x, y, z, t) = c^2 \Delta p_1(x, y, z, t) \tag{1.9}$$

avec $\Delta = \partial_{x^2} + \partial_{y^2} + \partial_{z^2}$, et l'équation de Helmholtz devient

$$\Delta\varphi(x, y, z) + k^2\varphi(x, y, z) = 0.$$

Par exemple, les *ondes harmoniques sphériques* provoquées par une source ponctuelle supposée placée à l'origine sont de la forme (avec $r = \sqrt{x^2 + y^2 + z^2}$) :

$$p_1(x, y, z, t) = \alpha \frac{\exp(ikr - 2i\pi f t)}{r}.$$

Ces ondes sont dites sphériques car, à t fixé, étant donnée une sphère centrée sur l'origine, la pression y est la même en tout point de cette sphère. À noter qu'elles ne sont pas stationnaires.

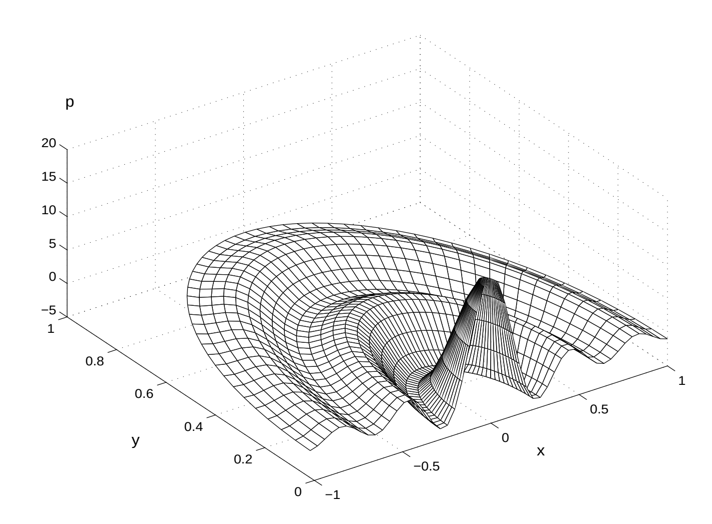


FIG. 1.7 – une onde sphérique harmonique (coupe 2D). Elle s'amortit en $1/r$.

1.1.4 Intensité sonore

Nous avons écrit plus haut la pression (cas d'une onde plane) sous la forme $p(x, t) = p_0 + \varepsilon p_1(x, t)$ où p_0 est la pression à l'équilibre, ou pression moyenne. La différence $p(x, t) - p_0$ est appelée *pression acoustique* p_a :

$$p_a(x, t) = p(x, t) - p_0.$$

Pour mettre l'air en mouvement, il a fallu fournir une certaine quantité de travail. La propagation de la déformation de l'air correspond à une propagation de l'énergie initiale. On

appelle *intensité sonore* le flux moyen d'énergie (en valeur absolue) par unité de temps et de surface. Elle est mesurée en Watts/m^2, et est donnée par la formule

$$I(x) = \frac{1}{T}\left|\int_0^T p_a(x,t)v(x,t)\,dt\right|$$

où l'échelle de temps T dépend du contexte. Cette intégrale peut être nulle si par exemple p_a et v sont en quadrature de phase (différence de phase égale à $\pi/2$). Dans le cas d'une onde plane progressive $p_a(x,t) = g(x-ct)$, le calcul *via* l'équation d'Euler (1.3) et l'équation d'état (1.5) donne $v(x,t) = p_a(x,t)/c\rho_0$, d'où

$$I(x) = \frac{1}{Tc\rho_0}\int_0^T p_a^2(x,t)dt. \tag{1.10}$$

Si l'onde est harmonique, par exemple $p_a(x,t) = p_\alpha \sin(kx - 2\pi ft)$, on obtient pour $T = 1/f$

$$I(x) = \frac{p_\alpha^2}{2c\rho_0} \simeq \frac{p_{\text{eff}}^2}{415},$$

avec $p_{\text{eff}} = p_\alpha/\sqrt{2}$, formule habituellement employée pour le calcul de l'intensité. Dans le cas d'une onde sphérique $p_a(x,y,z,t) = p_\alpha \sin(kr - 2\pi ft)/r$, le calcul donne

$$I(x,y,z) \simeq \frac{p_\alpha^2}{2c\rho_0} \times \frac{1}{r^2}.$$

L'intensité d'un son provenant d'une source ponctuelle est ainsi (en l'absence de dissipation) inversement proportionnelle au carré de la distance à cette source.

Le *seuil d'audition* est approximativement

$$I_0 = 10^{-12}\text{W/m}^2,$$

le niveau normal pour une conversation est de $1.2\,10^{-5}$W/m^2 et le *seuil de douleur* environ 1W/m^2. Nous verrons dans le chapitre 4 que ces seuils varient en fonction de différents paramètres, et tout particulièrement de la fréquence. On remarque en passant que la dynamique de l'oreille est assez impressionnante : 10^{12} ! À l'unité W/m^2, on préfère souvent le bel (du nom de A.G. Bell, professeur dans une école de malentendants et inventeur du téléphone) ou le *décibel*, unité a-dimensionnée qui mesure le logarithme en base 10 du rapport à un seuil donné, par exemple le seuil d'audition, ce qui donne en notant L_I cette nouvelle mesure de l'intensité acoustique :

$$L_I = 10\log\frac{I}{I_0}\text{ dB}.$$

Ainsi, le seuil d'audition est par convention de 0dB, le seuil de douleur est de 120dB, et celui d'une conversation est d'environ 70dB. Sachez que dans certains concerts de rock, l'intensité dépasse parfois les 140dB !

Une petite question : que se passe-t-il dans un orchestre symphonique lorsque l'on multiplie le nombre de violons par 10 ?

Réponse (cf. § 1.6.3) : le niveau sonore est augmenté de 10dB. Autrement dit, l'augmentation de volume paraît être la même lorsque l'on passe de 1 à 10 violons que lorsque l'on passe

de 10 à 100 violons ! C'est un cas particulier de la loi du physiologiste Fechner : *la sensation varie comme le logarithme de l'excitation* (cf. [10], mais aussi [15] pour un avis plus nuancé, que nous aborderons au chapitre 4). Le même phénomène se produit pour la perception de la hauteur : les intervalles [La-1, La-2] et [La-2, La-3] sont perçus comme égaux (une octave) alors qu'ils correspondent à un rapport 2 des fréquences : [110, 220] et [220, 440].

1.2 Un peu de solfège

Avant d'aller plus loin, il est peut-être nécessaire de nous remettre en mémoire quelques notions élémentaires de solfège et le vocabulaire associé. Une note de musique est caractérisée par trois paramètres principaux : sa durée, son intensité et sa hauteur. Nous nous intéressons ici à la hauteur. Celle-ci est reliée à la fréquence[3] de la note : aux basses fréquences correspondent les sons *graves*, et aux fréquences élevées correspondent les sons *aigus*. La fréquence de référence pour le musicien est le La à 440 Hz ou *diapason*, c'est la note que vous entendez lorsque vous décrochez votre téléphone.

1.2.1 Intervalles, octave

En théorie de la musique, la distance qui sépare deux notes distinctes est appelée *intervalle*. Lorsque notre oreille évalue l'intervalle entre deux notes, c'est le *rapport* de leur fréquence et non leur *différence* qui détermine notre perception de la distance entre ces deux notes. Nous retrouvons ici la loi de Fechner rencontrée à propos de l'intensité : la sensation de hauteur varie comme le logarithme de la fréquence (cette loi est en fait prise en défaut dans l'extrême grave et l'extrême aigu, nous y reviendrons au chapitre 4). Par exemple, les deux intervalles musicaux [110 Hz, 220 Hz] et [220 Hz, 440 Hz] sont perçus comme étant *égaux* car les rapports des fréquences sont égaux : $220/110 = 440/220$, alors qu'au sens mathématique du terme, le second intervalle est deux fois plus grand que le premier : $440 - 220 = 2 \times (220 - 110)$. L'intervalle entre deux telles notes est appelé *octave.*

1.2.2 Numérotation des notes

Les sons produits par deux notes à l'octave l'une de l'autre sont très ressemblants (nous verrons pourquoi au § 2.1.1), à tel point qu'ils sont désignés par la même note. Ainsi, la note de fréquence 880 Hz, à l'octave du La du diapason, produira aussi un La, mais plus aigu. Pour les distinguer entre elles, nous adopterons la convention suivante : le La à 440 Hz sera noté La-3, le suivant à 880 Hz sera noté La-4, suivi du La-5 à 1760 Hz etc. De même, on trouve en descendant le La-2 à 220 Hz, le La-1 à 110 Hz etc. Nous procèderons de la même manière pour les autres notes, en attribuant le suffixe 3 aux notes comprises entre le Do à 261.6 Hz et le Si à 493.9 Hz, et qui se trouvent à peu près au milieu du clavier du piano.

1.2.3 Partage de l'octave en douze demi-tons

Les autres intervalles sont déterminés par le choix du *tempérament*, qui fixe de manière précise les rapports de fréquence entre les différentes notes, et que nous discuterons en détail

[3] Nous verrons dans la section suivante qu'il s'agit de la fréquence du fondamental de la note.

au chapitre 3. Nous nous plaçons ici dans le cas du tempérament *égal.* Dans celui-ci, l'octave est partagée en douze intervalles égaux appelés *demi-tons,* de rapport de fréquence $2^{1/12} \simeq 1.0595$. Si partant d'une note de fréquence f, on monte douze fois de demi-ton en demi-ton, on obtient successivement les notes de fréquence $2^{1/12}f$, puis $2^{1/12}\times 2^{1/12}f = 2^{2/12}f$, puis $2^{3/12}f$,..., $2^{11/12}f$, et finalement $2^{12/12}f = 2f$, ce qui nous a bien amené à l'octave supérieure par intervalles égaux.

Ces treize notes forment ce que l'on appelle la *gamme chromatique,* qui fut inventée par les chinois il y a plus de quatre millénaires ! En partant de Do, ce sont les notes Do, Do♯, Ré, Ré♯, Mi, Fa, Fa♯, Sol, Sol♯, La, La♯, Si, Do, le signe ♯ indiquant que la note a été augmentée d'un demi-ton, la note obtenue étant dite *altérée.* En utilisant le signe ♭ pour abaisser une note d'un demi-ton, cette suite de notes peut également s'écrire Do, Ré♭, Ré, Mi♭, Mi, Fa, Sol♭, Sol, La♭, La, Si♭, Si, Do. Les notes Do♯ et Ré♭ sont dites *enharmoniques,* et sont égales dans le tempérament égal. Il en va de même pour les autres notes enharmoniques Ré♯ et Mi♭, Fa♯ et Sol♭ etc. À noter cependant que les musiciens qui, comme les violonistes, ont la possibilité de déterminer eux-mêmes la hauteur des notes, jouent très souvent le Do♯ légèrement plus haut que le Ré♭. L'intervalle composé de deux demi-tons s'appelle bien entendu un ton, il y en a donc six dans une octave. Les notes correspondantes forment la *gamme en tons* dont se servit abondamment Claude Debussy.

1.2.4 Gammes diatoniques

Les gammes classiques ne sont ni la gamme chromatique ni la gamme en tons, mais les gammes *diatoniques,* dans lesquelles on rencontre les deux types d'intervalles : le ton et le demi-ton, et qui constituent une suite de huit notes dont la dernière est à l'octave au-dessus de la première. Ces gammes résultent de la juxtaposition de deux *trétracordes* (quatre notes consécutives) comportant chacun obligatoirement deux tons, d'où le nom de diatonique. Le placement du demi-ton dans chaque tétracorde détermine alors les différents modes ou gammes possibles. Les notes d'une gamme sont appelées *degrés* de la gamme, le premier (et le huitième, puisque c'est la « même » note) s'appelant la *tonique,* le cinquième la *dominante* et le septième la *note sensible*, qui dans l'harmonie classique « appelle » à être suivie de la tonique.

Il y a plusieurs types de gammes diatoniques, dont deux ont joué un rôle central dans toute la musique classique : la gamme *majeure* et la gamme *mineure*, qui se distinguent essentiellement par leurs troisième et sixième degré, appelés pour cette raison notes *tonales.* Les airs enlevés, brillants ou gais (marches, airs de fête, danses) sont souvent écrits en majeur, alors que les airs recueillis, tristes ou sombres (requiem, nocturnes, marches funèbres) sont plutôt écrits en mineur. Les autres gammes sont appelées gammes *modales*, et furent largement utilisées tout au long du Moyen Âge, en particulier dans la musique ecclésiastique.

Gamme majeure

Une gamme diatonique majeure est constituée des intervalles suivants : ton, ton, demi-ton, ton, ton, ton, demi-ton. Par exemple, en partant de Do, cela donne la suite de notes Do, Ré, Mi, Fa, Sol, La, Si, Do, representées sur une portée de la manière suivante :
Ce procédé de notation est dû à Guy d'Arezzo (début du XIe siècle). Les différents degrés de la

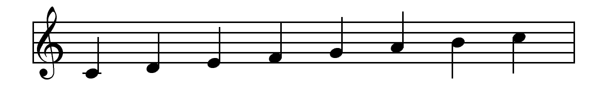


FIG. 1.8 – gamme en Do majeur, commençant par Do-3 et finissant par Do-4.

gamme sont alternativement posés *sur* et *entre* les lignes de la portée. Les intervalles Do–Ré, Do–Mi, Do–Fa,..., Do–Si s'appellent respectivement *seconde, tierce majeure, quarte, quinte, sixte majeure* et *septième majeure.* Cette appellation fait bien sûr référence à l'intervalle entre les notes et non aux notes qui le composent. Ainsi les intervalles Fa-La et Sol-Si constituent également une tierce majeure, composée de deux tons, et les intervalles Ré-La, Mi-Si et Fa-Do forment une quinte, composée de trois tons et demi et correspondant au rapport de fréquence $2^{7/12} \simeq 1.5$.

Gammes mineures

Il y deux types de gammes diatoniques mineures, utilisées selon que la mélodie monte ou descend :

- la gamme mineure *mélodique ascendante*, constituée des intervalles ton, demi-ton, ton, ton, ton, ton, demi-ton. En partant de Do, cela donne la suite de notes Do, Ré, Mi♭, Fa, Sol, La, Si, Do,

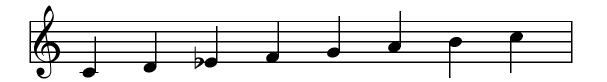


FIG. 1.9 – gamme en Do mineur mélodique ascendant.

- la gamme mineure *mélodique descendante*, constituée des intervalles ton, demi-ton, ton, ton, demi-ton, ton, ton. En prenant Do comme tonique, cela donne les notes Do, Ré, Mi♭, Fa, Sol, La♭, Si♭, Do. Si l'on prend La comme tonique, cela donne les notes La, Si, Do, Ré, Mi Fa, Sol, La, où aucune altération n'est présente. Cette dernière gamme est appelée gamme mineure *relative* à la gamme majeure en Do.

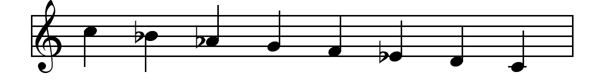


FIG. 1.10 – gamme en Do mineur mélodique descendant

Un peu à part, on trouve la gamme mineure *harmonique* (figure 1.11), utilisée comme son nom l'indique pour composer les accords (l'harmonie) accompagnant une mélodie écrite en mineur, et constituée des intervalles ton, demi-ton, ton, ton, demi-ton, un ton et demi, demi-ton. En partant de Do, cela donne la suite de notes Do, Ré, Mi♭, Fa, Sol, La♭, Si, Do.

Les intervalles Do–Mi♭, Do–La♭ et Do-Si♭ s'appellent respectivement *tierce mineure, sixte mineure* et *septième mineure,* et représentent des intervalles d'un ton et demi, quatre tons et cinq tons.

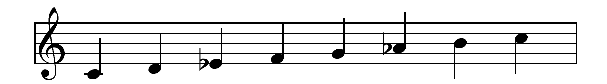


FIG. 1.11 – gamme en Do mineur harmonique

1.3 Quelques types de sons

Un auditeur placé en un point donné x de l'espace percevra la variation en ce point de la pression de l'air, qui est une fonction du temps que nous noterons $s(t) := p_a(x,t)$. C'est le *signal sonore.* Nous avons déjà rencontré le *signal sinusoïdal,* que l'on écrit sous différentes formes selon le contexte :

$$\begin{aligned} s(t) &= \alpha\cos(\omega t+\theta) \\ &= \alpha\cos(2\pi f t+\theta) \\ &= \mathrm{Re}(\alpha\exp(i\theta)\exp(2i\pi f t)) \\ &= a\cos(2\pi f t)+b\sin(2\pi f t) \\ &= c_1\exp(2i\pi f t)+c_2\exp(-2i\pi f t) \end{aligned}$$

où $\mathrm{Re}(z)$ désigne la partie réelle de z, et :

- $\omega \geq 0$ est la *pulsation* en radians/s,
- $f = \omega/2\pi \geq 0$ est la *fréquence*[4] d'unité le *hertz* (Hz) ; elle indique le nombre de vibrations par seconde,
- $\alpha \geq 0$ est *l'amplitude*,
- θ est la *phase* à l'origine, mesurée en radians avec $\theta \in [0, 2\pi[$,
- $a = \alpha\cos\theta$ (prendre $t = 0$), $b = \alpha\cos(\theta+\pi/2)$ (prendre $2\pi f t = \pi/2$),
- $c_1 = (a-ib)/2$, $c_2 = (a+ib)/2 = \overline{c_1}$ (utiliser $\exp(ix) = \cos x + i\sin x$).

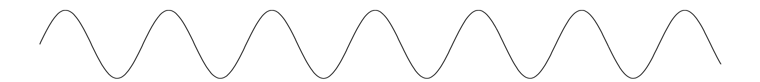


FIG. 1.12 – son sinusoïdal, qualifié de « son pur».

Ce son sinusoïdal est l'un des sons les plus simples, on dit que c'est un *son pur.* Pour une onde plane $p_a(x,t) = \alpha\cos(kx+2\pi f t)$, c'est celui qui se produit en tous les points x tels que $kx = \theta + 2n\pi$, $n \in \mathbb{Z}$.

Une propriété importante de l'équation des ondes est d'être linéaire et homogène. Ceci entraîne que si $p_1(x,t) = \cos(k_1 x + 2\pi f_1 t)$ et $p_2(x,t) = \cos(k_2 x + 2\pi f_2 t)$ sont solutions de cette équation (ce qui est le cas pour $k_i = 2\pi f_i/c$, $i = 1,2$), alors $\alpha_1 p_1(x,t) + \alpha_2 p_2(x,t)$ sera aussi solution de l'équation des ondes. En un point donné x, le son perçu sera alors de la forme $s(t) = \alpha_1\cos(2\pi f_1 t + \theta_1) + \alpha_2\cos(2\pi f_2 t + \theta_2)$. Ce son, plus complexe, est la superposition

[4] Pour une fonction réelle, la fréquence est toujours supposée positive ou nulle. En écriture complexe, sachant que $\cos(2\pi f t) = [\exp(2\pi i f t) + \exp(-2\pi i f t)]/2$, on est obligé de considérer également le cas des fréquences négatives (ici $-f$).

des deux fréquences f_1 et f_2. En continuant ce procédé, on constate qu'une onde acoustique peut produire en un point de l'espace un signal sonore de la forme

$$s(t) = \sum_{n \geq 1} \alpha_n \cos(2\pi f_n t + \theta_n). \quad (1.11)$$

Si la somme comporte un nombre infini de termes, certaines conditions sont à imposer sur les α_n et les f_n pour que la somme converge. Le spectre d'un tel son, c'est-à-dire l'ensemble des fréquences f_n présentes, est dit *spectre discret.*

On considère qu'une excellente oreille peut percevoir les fréquences situées entre 20Hz et 20 kHz et que les sons deviennent inaudibles en dehors (infra-sons ou ultra-sons).

1.3.1 Sons périodiques

Un cas intéressant se produit lorsque toutes les fréquences sont des multiples *entiers* d'une fréquence donnée $f > 0$: $f_n = nf$. Dans ce cas, le signal

$$s(t) = \sum_{n \geq 1} \alpha_n \cos(2\pi n f t + \theta_n)$$

est périodique de période $T = 1/f$, c'est-à-dire que $s(t + T) = s(t)$ pour tout t. En effet,

$$\cos(2\pi n f(t + T) + \theta_n) = \cos(2\pi n f t + 2n\pi + \theta_n) = \cos(2\pi n f t + \theta_n).$$

Bien entendu, il n'existe pas de sons réellement périodiques, ne serait-ce que parce qu'ils n'ont pu commencer avant le big-bang !

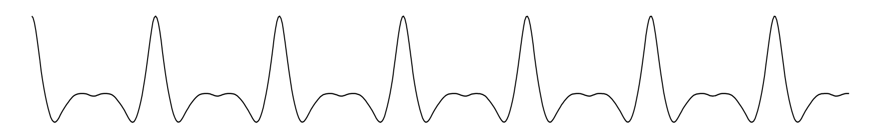


FIG. 1.13 – son périodique (approximation d'un son de trompette).

En musique, la fréquence f est appelée le *fondamental*, et la fréquence $f_n = nf$ est appelée *l'harmonique de rang* n (à ne pas confondre avec une onde *harmonique*). Si par exemple f est la fréquence du Do-3 (261.6Hz), alors f_2 est à *l'octave* au-dessus (Do-4), f_3 se situe une *quinte* plus haut (Sol-4), f_4 est l'octave suivante (Do-5), f_5 est légèrement en-dessous de la *tierce* majeure Mi-5, etc.

$f = 261.6$Hz	$2f$	$3f$	$4f$	$5f$	$6f$	$7f$	$8f$	$9f$
Do	Do	Sol	Do	Mi	Sol	Si♭	Do	Ré

(1.12)

Un son comportant un grand nombre d'harmoniques sera perçu comme « riche» (exemple du clavecin ou du violon), alors qu'un son ne comportant que peu d'harmoniques sera perçu comme « pauvre» (exemple de la flûte douce).

La figure 1.14 montre le signal obtenu en ajoutant un par un les termes $(\sin 2\pi nt)/n$, $n = 1, 2, 3, 4$. On s'approche de plus en plus d'un signal *triangulaire*, qui est le modèle le plus simple d'un son de violon.

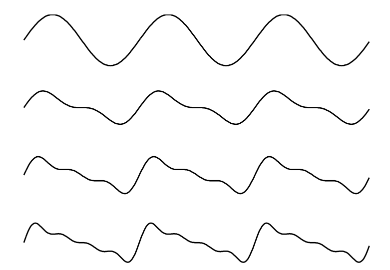


FIG. 1.14 – ajout des harmoniques 1, 2, 3, 4 (modèle simplifié d'un son de violon).

La figure 1.15 est obtenue de la même manière, mais avec les harmoniques de rang impair $n = 1, 3, 5, 7$. Ici on s'approche d'un signal *rectangulaire*, qui est le modèle le plus simple d'un son de clarinette.

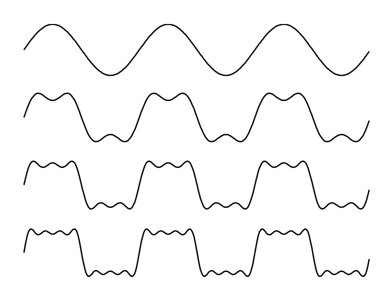


FIG. 1.15 – ajout des harmoniques 1, 3, 5, 7 (modèle simplifié d'un son de clarinette).

Typiquement, les instruments à *son entretenu* comme le violon, l'orgue ou l'accordéon produisent un son périodique, au moins sur une période de temps significative. L'analyse de tels sons peut se faire avec l'outil mathématique suivant.

Séries de Fourier

Le mathématicien Joseph Fourier (1768-1830) est à l'origine de l'analyse des sons périodiques et de leur décomposition en séries trigonométriques qui portent son nom : les séries de Fourier. Cette théorie a été élaborée alors qu'il étudiait la propagation de la chaleur dans un solide.

Si $s(t)$ est un son T-périodique, intégrable sur l'intervalle $[0, T]$, les *coefficients de Fourier* c_n, $n \in \mathbb{Z}$, sont définis par

$$c_n = \frac{1}{T} \int_0^T s(t) \exp(-2i\pi nt/T)\, dt. \tag{1.13}$$

Ceci constitue l'*analyse* de Fourier. On montre alors que sous certaines conditions supplémentaires,

la série ci-dessous, appelée *série de Fourier*, converge vers $s(t)$:

$$s(t) = \sum_{n=-\infty}^{+\infty} c_n \exp(2i\pi nt/T).$$

Ceci constitue la *synthèse* de Fourier : le signal $s(t)$ est reconstitué à partir de la somme de ses composantes fréquentielles $c_n \exp(2i\pi nt/T)$ de fréquence n/T. Cette somme peut également s'écrire avec des sinus et cosinus :

$$s(t) = \frac{a_0}{2} + \sum_{n=1}^{+\infty} a_n \cos(2\pi nt/T) + b_n \sin(2\pi nt/T)$$

avec $a_n = c_n + c_{-n}$ et $b_n = i(c_n - c_{-n})$.

Lorsque $s^2(t)$ est intégrable sur l'intervalle $[0, T]$, l'intensité du son sur une période (cf. (1.10)) est reliée aux coefficients de Fourier par la relation de Parseval

$$\frac{1}{T}\int_0^T |s(t)|^2\, dt = \sum_{n=-\infty}^{+\infty} |c_n|^2.$$

1.3.2 Sons avec partiels

D'autres instruments, comme par exemple les cloches, la plupart des instruments à percussion (fig. 1.16), ainsi que le piano dans une faible mesure, produisent des sons de la forme (1.11), mais qui ne sont plus périodiques. En termes de fréquences, cela se traduit par le fait qu'il n'existe pas de fréquence f telle que toutes les fréquences f_n soient des multiples entiers de f. Dans un tel cas, l'analyse par série de Fourier ne s'applique plus directement.

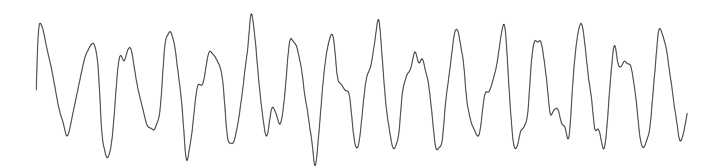


FIG. 1.16 – exemple de son non périodique (timbale).

Les fréquences présentes dans le son sont alors appelées *partiels*. Un exemple de son de cloche donné dans [13] est composé des fréquences[5] $0.5f_p$, f_p, $1.2f_p$, $1.5f_p$, $2f_p$, $2.5f_p$ et $3f_p$. Le deuxième partiel est appelé le *principal*, c'est lui qui donne la hauteur de la note. Les fondeurs de cloches accordent souvent celles-ci pour que le troisième partiel soit situé une tierce mineure (de rapport $2^{3/12} \simeq 1.189$) au-dessus du principal, ce qui leur confère leur sonorité caractéristique. On trouve ensuite une quinte, une octave, une tierce majeure... En

[5]En toute rigueur, ce son est aussi périodique, de période $10/f_p$, les 7 fréquences étant des multiples entiers de $f_p/10$. Mais en le considérant comme tel, les 4 premiers harmoniques sont nuls. De plus, si on ajoute des harmoniques plus élevés, la valeur $f_p/10$ ne conviendra plus, il faudra la diminuer. Enfin, ces fréquences ne sont que des approximations, il se peut très bien que pour les valeurs exactes on ne puisse trouver aucun diviseur commun.

fait, l'exemple précédent ne contient pas assez de partiels pour reconstruire un son réaliste. L'analyse d'une grande cloche a donné, pour $f_p = 233$Hz (Si♭-2), les fréquences suivantes et leur intensités respectives :

$$f = [0.5,\ 1,\ 1.188,\ 1.530,\ 2.000,\ 2.470,\ 2.607,\ 2.650,\ 2.991,\ 3.367,\ 4.137,\ 4.487,\ 4.829,\ 5.385,\ 5.863,\ 6.709,\ 8.077,\ 8.547,\ 9.017,\ 9.530,\ 11.026,\ 12.393] \times f_p,$$

$$I = [350,\ 950,\ 500,\ 150,\ 700,\ 100,\ 250,\ 370,\ 1000,\ 180,\ 300,\ 100,\ 150,\ 300,\ 100,\ 100,\ 50,\ 20,\ 10,\ 35,\ 5,\ 15].$$

Combinées avec une enveloppe adéquate (cf. § 2.5.2), ces données donnent une reconstitution assez ressemblante.

1.3.3 Sons à spectre continu

Le chant est encore assez proche d'un son périodique, mais avec des sons plus complexes comme la voix parlée, non seulement on s'éloigne des sons périodiques, mais on n'a même plus affaire à une somme de fréquences *ponctuelles* comme dans (1.11). On se trouve au contraire en présence d'un ensemble *continu* de fréquences : au lieu de s'écrire comme une *somme* de termes en $\exp(2i\pi f_n t)$ de fréquences f_n, de tels sons $s(t)$ peuvent se représenter à l'aide d'une *intégrale* de termes de même nature (synthèse de Fourier)

$$s(t) = \int_{-\infty}^{+\infty} \widehat{s}(f) \exp(2i\pi f t)\, df \tag{1.14}$$

où la fonction $\widehat{s}(f)$, appelée *transformée de Fourier* de s, est donnée par la formule d'analyse

$$\widehat{s}(f) = \int_{-\infty}^{+\infty} s(t) \exp(-2i\pi f t)\, dt. \tag{1.15}$$

Chaque nombre complexe $\widehat{s}(f)$ indique l'amplitude et la phase à la fréquence f. Le spectre d'un tel son, par opposition au spectre discret dans (1.11), est dit *spectre continu.* Cette représentation s'applique en particulier à toute fonction $s(t)$ intégrable sur $\mathbb{R}$ telle que $\widehat{s}(f)$ soit aussi intégrable sur $\mathbb{R}$.

Cependant, une telle représentation ne donne pas entière satisfaction (pas plus d'ailleurs que (1.11)), car la fonction $\widehat{s}(f)$ ne donne pas d'information directe sur le déroulement temporel du signal sonore. On souhaiterait pouvoir utiliser une autre représentation faisant intervenir à la fois temps et fréquence, comme celle qui est utilisée par les musiciens lorsqu'ils retranscrivent sur une partition un morceau de musique. Par exemple, les figures 1.17 et 1.18 représentent l'analyse du chant de Maria Callas dans *la Norma* de Bellini. Nous expliquerons dans la section suivante comment est obtenue une telle représentation. Vous remarquerez le célèbre *vibrato* qui a suscité tant d'admirateurs, et sa remarquable régularité !

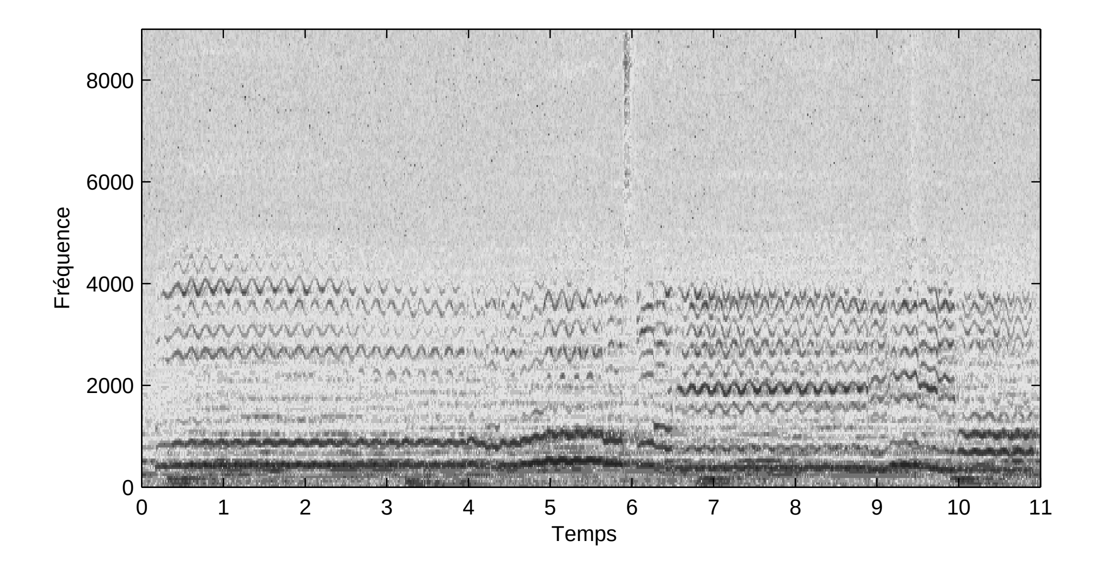


FIG. 1.17 – début de *la Norma* par Maria Callas. La première note est un La-3 ($441Hz$). Les deux premiers harmoniques sont très nets. Il y a ensuite un « trou» entre 1000 et 2500Hz, et l'énergie fréquentielle réapparaît entre 2500 et 4000Hz, là où elle est particulièrement efficace pour « couvrir»l'orchestre. A noter bien sûr le vibrato, mais aussi le « s» à $t = 6$.

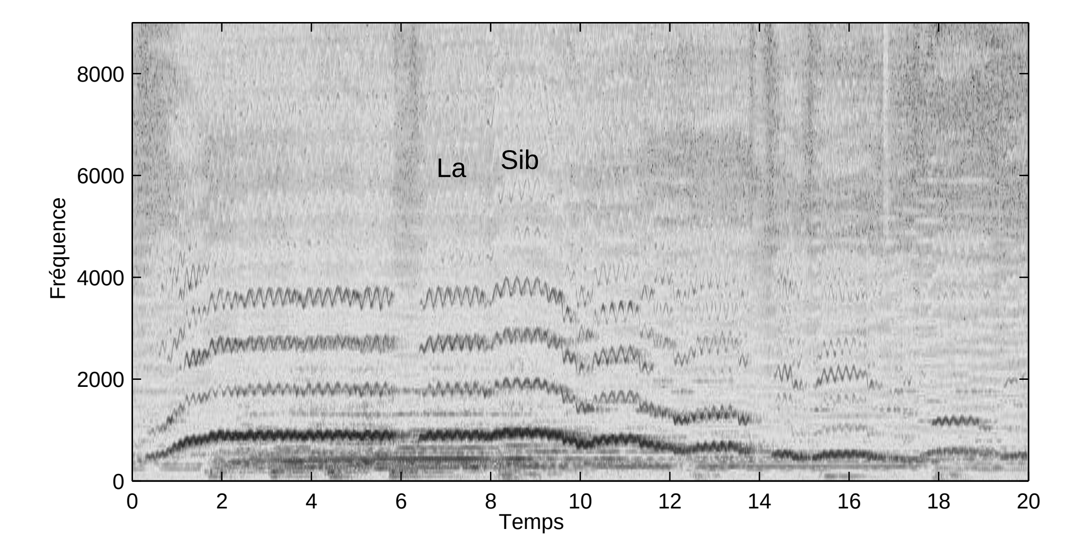


FIG. 1.18 – plus loin, la célèbre montée « Si Do Ré Mi Fa Sol La, La Si♭, La Sol Fa Sol...»

Enfin, les humains n'étant pas les seuls à chanter, voici pour les amoureux de la nature le chant du chardonneret de nos campagnes, en temps et temps-fréquence...

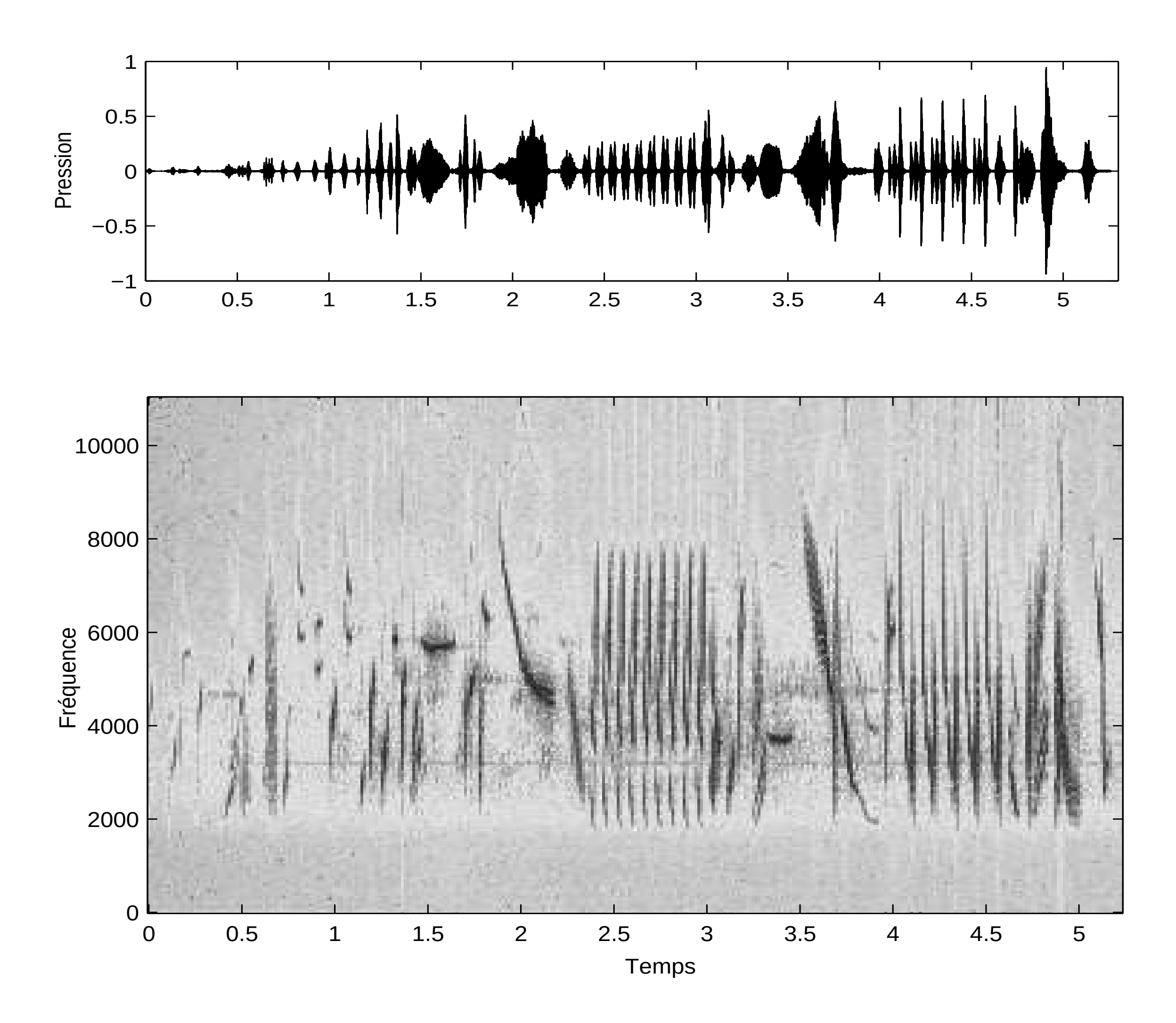


FIG. 1.19 – une petite touche bucolique, dont nous régale le gourmand de chardons. Cela pourrait être un tableau... musical !

ou, pour ceux qui préfèrent l'altitude, celui du siffleur des montagnes. Au passage, on remarquera la manière dont le signal temporel est « enveloppé», nous en reparlerons à propos des instruments au § 2.5.2.

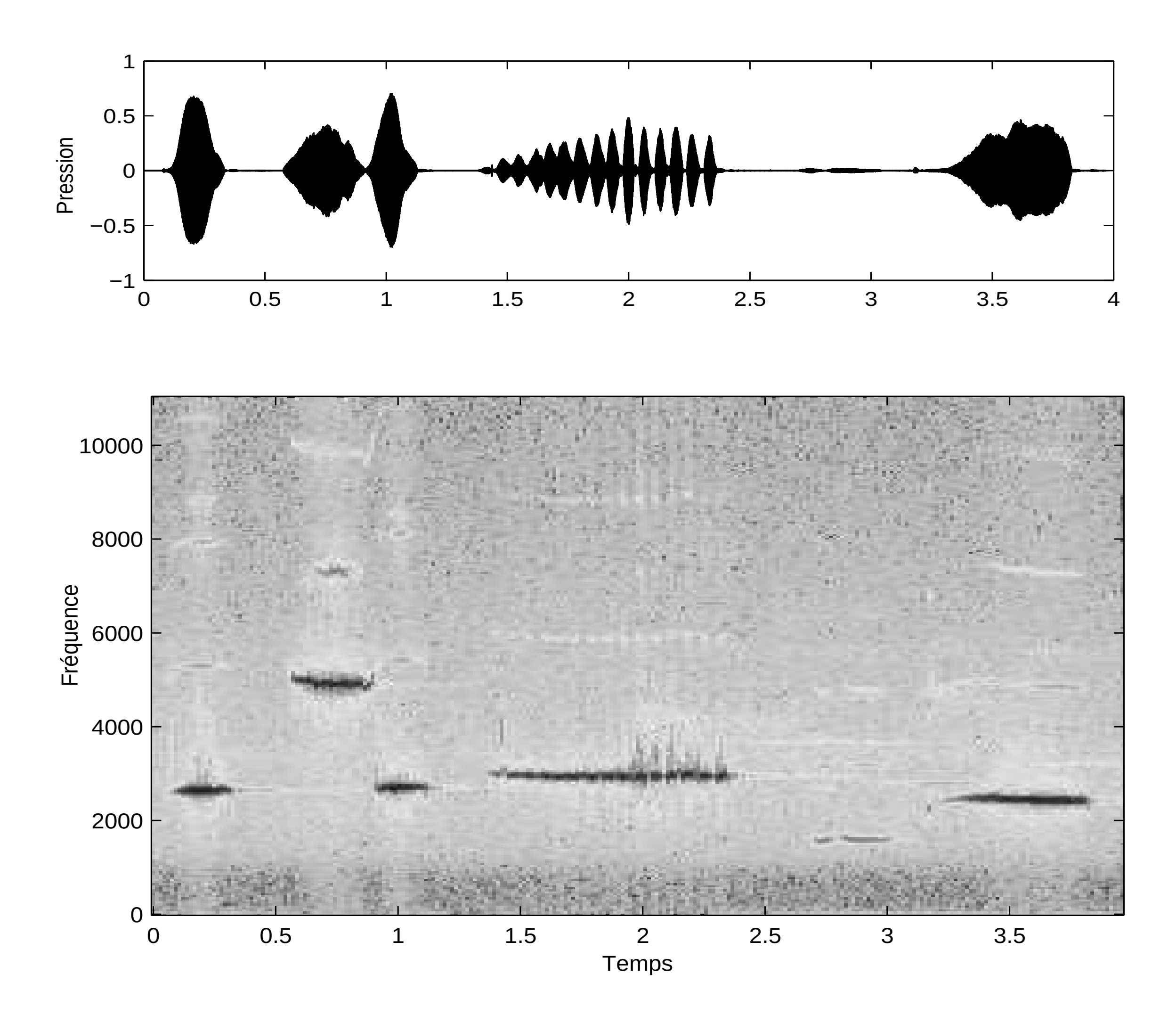


Fig. 1.20 – le siffleur des montagnes : peut-être moins pictural, mais beaucoup plus mélodique. L'air qui nous est interprété est très proche de Mi Do Mi Fa♯ Sol Ré.

1.3.4 Le bruit

Fin de la séquence nature, revenons dans nos villes. Par quoi sommes-nous entourés en permanence ? par du bruit ! Le bruit, certes moins charmant, est associé à la notion de son non structuré, présentant un caractère aléatoire. Par exemple, un signal $s(t)$ où chaque valeur $s(t)$ est une variable aléatoire suivant une loi donnée (gaussienne par exemple), les variables $s(t)$ et $s(t')$ étant indépendantes deux à deux pour tout $t \neq t'$, est un bruit qui ressemble à un souffle continu, ou à la friture qui apparaît lorsqu'un récepteur radio n'est branché sur aucune station.

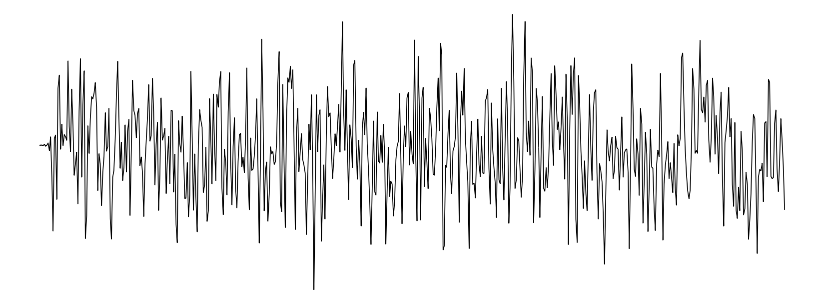


FIG. 1.21 – vous n'avez pas entendu un bruit ?

Le *bruit blanc* (par analogie avec la lumière blanche) est un son dans lequel toutes les fréquences audibles sont également présentes. Par exemple, le son suivant :

$$s(t) = \sum_{n=1}^{N} \cos(2\pi f_n t + \theta_n)$$

fera un bruit blanc tout à fait convenable pour N est assez grand (quelques centaines), les fréquences f_n étant choisies de manière aléatoire et uniforme sur l'intervalle $[20\text{Hz}, 20\,\text{kHz}]$ et les phases θ_n choisies de la même manière sur l'intervalle $[0,\, 2\pi]$. Dans le *bruit rose,* c'est $\log f_n$ qui est choisi aléatoirement et uniformément sur $[\log 20,\, \log 20000]$, de sorte que la puissance (l'intensité) est la même à l'intérieur de toute octave.

Le bruit blanc est par exemple utilisé par les ingénieurs du son pour repérer avant le concert les fréquences de résonance de la salle : ces fréquences seront moins atténuées que les autres dans la réponse de la salle au bruit blanc. Ceci signifie qu'au cours du concert, ces fréquences seront amplifiées par rapport aux autres. S'il s'agit d'un concert sonorisé, l'ingénieur du son pourra alors compenser cet effet indésirable en utilisant l'équaliseur de sa table de mixage : il lui suffira d'atténuer avec des *filtres* (cf. § 1.5) adéquats ces fréquences de résonance. En filtrant un bruit blanc avec des filtres *passe-bande* (qui ne laissent passer que les fréquences situées dans un intervalle donné), on peut obtenir toute une gamme de bruits et les utiliser pour des effets sonores, par exemple pour imiter le *souffle* produit par un instrument à vent.

1.4 Représentation du son

Lors de l'étude des quelques types de sons des paragraphes précédents, nous avons été confrontés au problème de la représentation du son : sous forme temporelle ou fréquentielle. Chacune a ses avantages et ses inconvénients, mais si l'on suit l'intuition musicale, on sent la nécessité d'une représentation faisant intervenir à la fois le temps et la fréquence : c'est ce que l'on appelle une *représentation temps-fréquence*. C'est un domaine où la théorie est à la fois riche et complexe. Nous nous contenterons d'en décrire une version simplifiée, en l'introduisant par un exemple musical. Supposons que l'on joue à la flûte la suite de notes La, Do, Mi de fréquences 440, 523.25 et 660Hz, chacune d'une durée d'une seconde, en commençant à l'instant $t = 0$. Le musicien écrira :
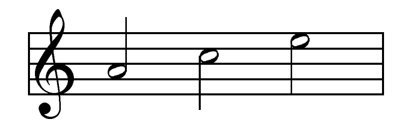


FIG. 1.22 – représentation musicale de la séquence La - Do - Mi.

On suppose que chaque note contient les quatre premiers harmoniques, d'amplitudes respectives 64, 16, 4 et 1 (les α_n). Effectuons alors sur chaque intervalle de temps $[0, 1]$, $[1, 2]$ et $[2, 3]$ l'analyse des coefficients de Fourier, et reportons sur une figure ces coefficients en fonction du temps, avec un trait dont l'épaisseur varie avec l'amplitude. On obtient alors le graphique de la figure 1.23, appelé *sonagramme* ou *spectrogramme*. Dans ce graphique, la

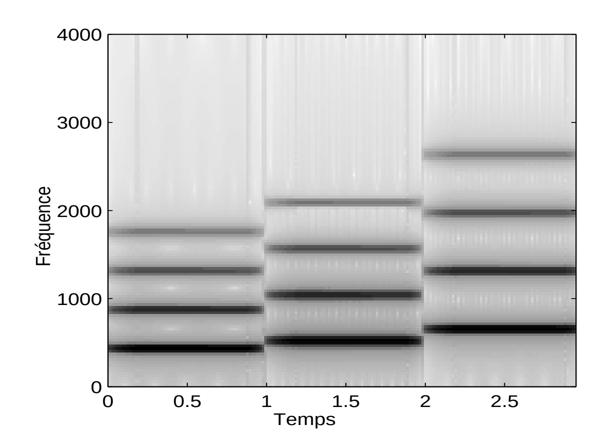


FIG. 1.23 – spectrogramme de La-Do-Mi, dans le cas de quatre harmoniques par note.

fréquence est en ordonnée et le temps en abscisse. Le parallèle avec la partition (fig. 1.22) est évident, le spectrogramme comportant l'information supplémentaire sur la composition en harmoniques du son analysé. En fait, un tel spectrogramme n'est pas calculé en prenant des intervalles d'une seconde, qui seraient beaucoup trop longs pour des sons variant plus rapidement que celui de l'exemple précédent. Nous décrirons un peu plus loin le procédé exact utilisé.

1.4.1 Transformée de Fourier Discrète

Les trois figures 1.24, 1.25 et 1.27 montrent trois types d'analyse du son parlé enregistré « le chapeau». La figure 1.24 représente le signal temporel. La figure 1.25 montre l'analyse en

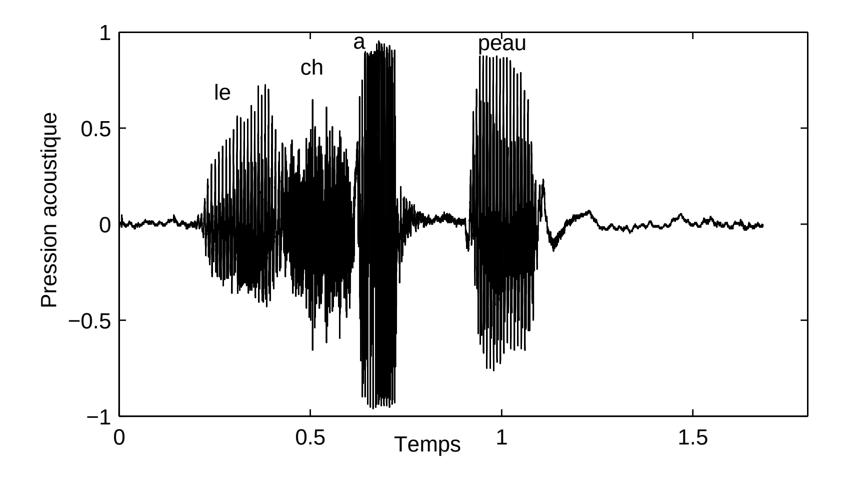


Fig. 1.24 – représentation temporelle de « le chapeau», assaisonnée d'un léger bruit de fond.

fréquence de ce même signal. Pour cela on a considéré que le signal enregistré $s(t)$ représentait

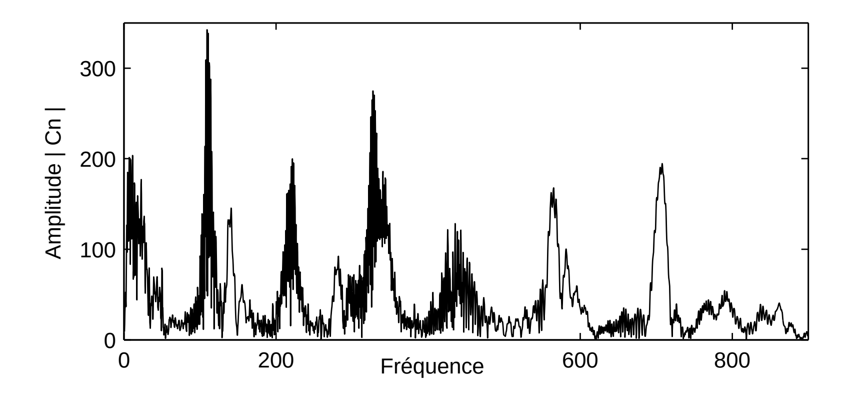


Fig. 1.25 – représentation fréquentielle de « le chapeau». Les voyelles sont responsables des pics aux multiples de 110Hz.

une période (de longueur $T = 1.7$s) d'un signal périodique dont on a calculé les coefficients de Fourier c_n :

$$s(t) = \sum_{n=-\infty}^{+\infty} c_n \exp(2i\pi nt/T) \text{ pour } 0 \leq t \leq T.$$

En fait, ces coefficients c_n ne sont pas calculés de manière exacte par la formule (1.13), mais par une intégration approchée utilisant la formule des trapèzes à des instants $t_k = kT/N$,

$k = 0, 1, \ldots, N-1$, qui donne

$$c_n \simeq \widetilde{c}_n := \frac{1}{N} \sum_{k=0}^{N-1} s(t_k) \exp(-2i\pi nk/N). \tag{1.16}$$

C'est ce que l'on appelle la *transformée de Fourier discrète* TFD (DFT en anglais) d'ordre N. Comme pour la série de Fourier, une transformation inverse permet de retrouver les $s(t_k)$ à partir des $\widetilde{c}_n$:

$$s(t_k) = \sum_{n=0}^{N-1} \widetilde{c}_n \exp(2i\pi nk/N). \tag{1.17}$$

La fréquence la plus basse[6] est donc ici $f_1 = 1/1.7$Hz. Les $\widetilde{c}_n$ (en fait leur amplitude) sont représentés sur la figure 1.25 non pas en fonction de n (c'est un entier qui n'a aucun sens physique) mais en fonction de la fréquence associée $f_n = n/1.7$Hz. On constate que certaines fréquences sont plus fortes que d'autres, en particulier autour des fréquences 110Hz, 220Hz, 330Hz, 440Hz, 570Hz, 700Hz qui correspondent à peu près aux harmoniques d'une note de fondamental 110Hz (La-1) produite par les *cordes vocales*. Le défaut de cette représentation (fig. 1.25) est de ne donner aucune information sur l'évolution temporelle du signal sonore, alors que celui de la figure 1.24 était de ne donner aucune information sur les fréquences présentes dans ce son. D'où l'intérêt d'introduire un autre type d'analyse : l'*analyse temps-fréquence* fournie par le spectrogramme.

1.4.2 Le spectrogramme

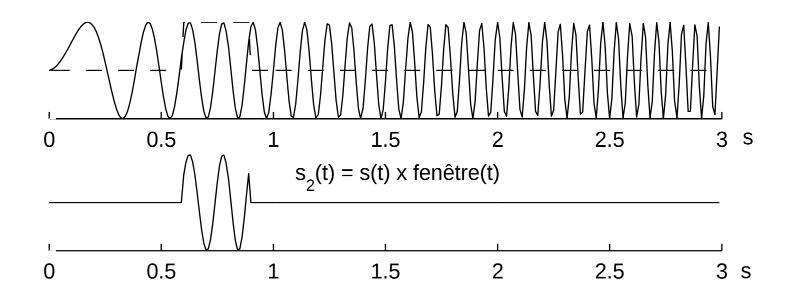


FIG. 1.26 – multiplication du signal par une fenêtre rectangulaire.

Le principe pour effectuer l'analyse temps-fréquence d'un signal sonore $s(t)$, $t \in [0, T]$, est le suivant. On commence par découper le signal en M petites tranches $s_k(t)$ telles que

$$s_k(t) = \begin{cases} s(t) \text{ si } t \in I_k = [kT/M,\, (k+1)T/M[, \\ 0 \text{ sinon.} \end{cases}$$

Une autre façon d'écrire cela est de poser

$$s_k(t) = w(Mt/T - k)\, s(t)$$

[6]Cette fréquence n'a aucun sens physique, et ne fait que refléter le fait que le son a une durée de 1.7s.

avec

$$w(t) = \begin{cases} 1 \text{ si } t \in [0,\, 1[, \\ 0 \text{ sinon,} \end{cases}$$

c'est-à-dire que chaque tranche $s_k(t)$ est obtenue en multipliant $s(t)$ par la *fenêtre glissante* rectangulaire $w(Mt/T - k)$. La figure 1.26 montre un signal s et la troisième tranche s_2 pour $T = 3$ et $M = 10$. Remarquez ici comme la fréquence augmente avec le temps dans le signal s, comme au début d'une sirène.

Sur chaque intervalle I_k, $k = 0, 1, \ldots, M-1$, on effectue ensuite une TFD du signal $s_k(t)$, et, pour chaque coefficient c_n^k (l'indice k indiquant que l'on est sur l'intervalle I_k) associé à la fréquence $f_n = nM/T$, on reporte sur le spectrogramme un trait joignant les points de coordonnées $(kT/M,\, f_n)$ et $((k+1)T/M),\, f_n)$ avec une épaisseur proportionnelle à $|c_n^k|$. Typiquement, chaque TFD est d'ordre 256 ou 512. Dans la pratique, au lieu d'utiliser des intervalles disjoints (qui peuvent rendre « invisible» ce qui se passe à la jonction, un saut par exemple), on utilise des intervalles qui se chevauchent, et au lieu d'une fenêtre rectangulaire, on utilise des fenêtres qui ont une coupure moins brusque, comme celles représentées sur la figure 5.10.

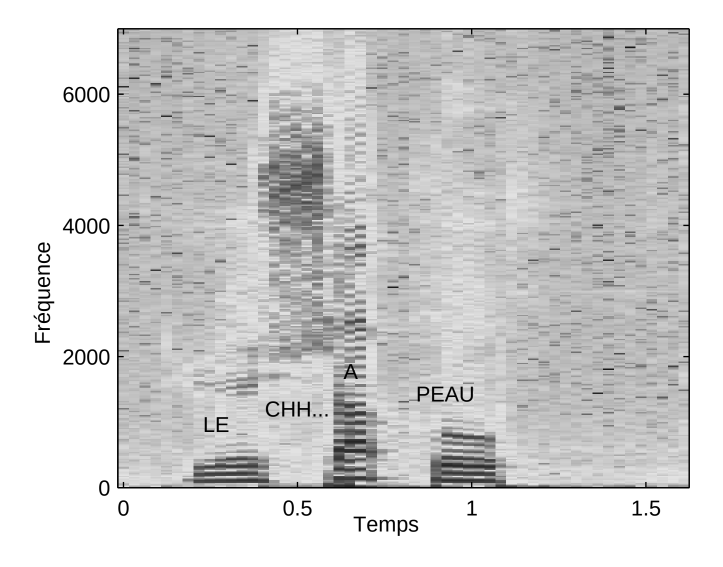


FIG. 1.27 – représentation temps-fréquence de « le chapeau». Les voyelles sont assez « musicales», la consonne *ch* est plutôt « bruyante» !

Appliqué à notre exemple, ce procédé donne la figure 1.27. On constate en particulier qu'une consonne comme « ch» est bien plus riche en fréquences élevées qu'une voyelle. On voit également apparaître dans les sons « e», « a» et « o» les harmoniques que nous avions observés dans la figure 1.25. L'évolution au cours du temps du signal reste cependant présente, nous avons bien à la fois des informations temporelles et fréquentielles.

Toutefois, on constate que cette image manque de netteté. Cela n'est pas dû à un problème technique de traitement de l'image, mais à un réelle impossibilité d'avoir un signal concentré à la fois en temps et en fréquence, et cela pour deux raisons.

La première raison, qualitative, est que pour toute fonction donnée $s(t)$, il est impossible que les fonctions $s(t)$ et $\widehat{s}(f)$ soient toutes deux *à support compact*, c'est-à-dire nulles en dehors d'un intervalle borné (mis à part bien sûr le cas $s = 0$). En particulier, si un son a une durée finie, alors il contient nécessairement des fréquences arbitrairement élevées : quel que soit le seuil choisi f_S, ce son possède des composantes fréquentielles de coefficient non nul et de fréquence supérieure à f_S !

La seconde raison, quantitative, est connue en mécanique quantique comme le *principe d'incertitude de Heisenberg,* qui énonce qu'il est impossible de mesurer avec une précision arbitraire à la fois la position et la vitesse d'une particule. Ce principe est en fait un résultat mathématique qui dit que pour une fonction $s(t)$ de carré intégrable telle que $\int |s(t)|^2 dt = 1$, l'écart type σ associé à la densité de probabilité $|s(t)|^2$ et l'écart type $\widehat{\sigma}$ associé à la densité de probabilité $|\widehat{s}(t)|^2$ obéissent à l'inégalité

$$\sigma \, \widehat{\sigma} \;\geq\; \frac{1}{4\pi}. \tag{1.18}$$

Quelle est la conséquence de ces deux propriétés ? Imaginons que dans le graphique de notre spectrogramme, nous n'ayons qu'un petit carré noir de dimension $\varepsilon \times \varepsilon$ correspondant à un tel signal $s(t)$. Cela signifierait que la durée du signal dans le temps est ε, et que les fréquences seraient concentrées dans un intervalle de longueur ε, ce qui contredit déjà la première propriété. Mais même si cela était possible, on aurait nécessairement $\sigma \leq \varepsilon/2$ ainsi que $\widehat{\sigma} \leq \varepsilon/2$, ce qui reporté dans (1.18) donne $\varepsilon^2 \geq 1/\pi$, d'où $\varepsilon \geq 1/\sqrt{\pi}$, résolution graphique en-dessous de laquelle il est impossible de descendre dans la représentation temps-fréquence !

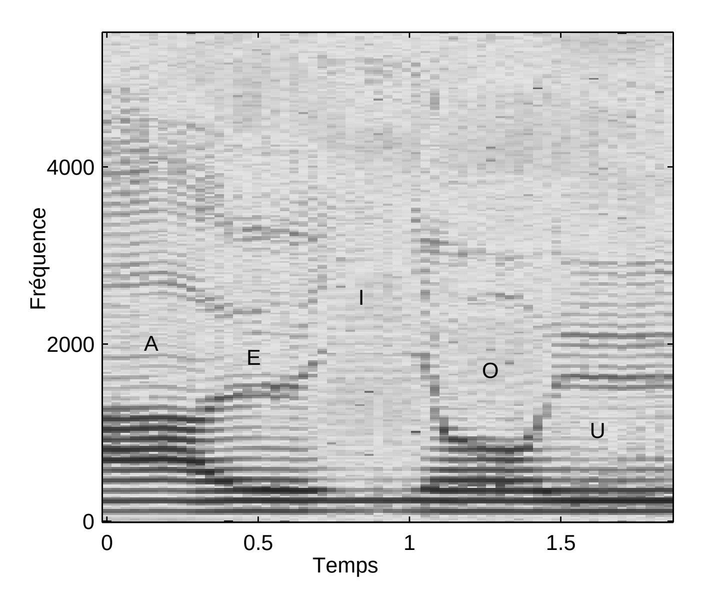


FIG. 1.28 – représentation temps-fréquence des voyelles. Regardez le son « o» : comme tout « o» qui se respecte, il fait un joli rond !

Terminons par un dernier exemple de spectrogramme : l'analyse des 5 voyelles (fig. 1.28). Ces sons sont plutôt concentrés sur les basses fréquences. Par ailleurs, pour chaque voyelle, on

voit apparaître des zones plus foncées qui correspondent aux différentes résonances produites par les variations de forme de l'appareil vocal : on les appelle les *formants.* Avec un peu d'attention, ces formants sont également visibles sur la figure 1.17, où l'on constate que pendant les six premières secondes, les harmoniques supérieurs ne suivent pas du tout la ligne mélodique (La Sol La Do Si♭ La Sol Sol) des deux premiers, mais semblent au contraire faire du « sur place». Cela provient du fait que ce ne sont pas les harmoniques aigus de *même rang* (au début 1, 2, 6, 7, 8, 9) qui sont intenses dans chacune de ces notes, mais *ceux qui se trouvent dans la plage 2500 à 4000Hz*. Le son est *mis en forme* par l'appareil vocal, qui renforce certaines fréquences et en atténue d'autres, d'où le nom de formants.

1.5 Filtrage

Le filtrage d'un son correspond entre autres à l'opération que vous effectuez lorsque vous tournez le bouton des aigus ou des graves sur votre chaîne audio, ce qui a pour effet de modifier la proportion graves/aigus du son. Nous en donnons ici une brève description mathématique.

1.5.1 Cas du spectre discret

Revenons pour commencer à un son pur, ne comportant donc qu'un seul harmonique de fréquence f, écrit sous forme complexe :

$$s(t) = \alpha \exp(2i\pi f t).$$

On peut effectuer sur ce son deux opérations de base qui sont :

- *l'amplification* d'un facteur $a > 0$:

$$v_1(t) = as(t),$$

- le *déphasage* d'angle $\theta \in [0, 2\pi[$:

$$v_2(t) = \alpha \exp(2i\pi f t - i\theta).$$

  Ce déphasage peut être aussi interprété comme un *retard* de $\tau = \theta/2\pi f$ puisque $2\pi f t - \theta = 2\pi f(t - \theta/2\pi f)$, et par conséquent

$$v_2(t) = s(t - \tau).$$

Si l'on effectue les deux opérations simultanément, on obtient le son

$$v(t) = a\alpha \exp(2i\pi f(t - \tau)) = a \exp(-2i\pi f\tau)s(t),$$

et le son initial a été multiplié par le nombre complexe

$$H = a \exp(-2i\pi f\tau),$$
$$v(t) = Hs(t).$$

Cette opération constitue la base du *filtrage :* amplification et/ou déphasage. Cette opération est évidemment *linéaire*, et *invariante par translation du temps*, c'est-à-dire que filtrer puis

translater d'un intervalle de temps t_0 donne le même résultat que translater de ce même intervalle de temps puis filtrer :

$$\begin{array}{llll} s(t) \longmapsto & u_1(t) = Hs(t) & \mapsto u_2(t) = u_1(t - t_0) \\ s(t) \longmapsto & v_1(t) = s(t - t_0) & \mapsto v_2(t) = Hv_1(t). \end{array} \tag{1.19}$$

On vérifie en effet que l'on a $u_2(t) = v_2(t) = Hs(t - t_0)$.

Fonction de transfert

Prenons maintenant un son plus complexe,

$$s(t) = \sum_n c_n \exp(2i\pi f_n t),$$

et appliquons à chaque terme de fréquence f_n l'opération précédente de multiplication par un nombre complexe dépendant de la fréquence, noté $H(f_n)$. On obtient en sortie un son modifié

$$v(t) = \sum_n H(f_n) c_n \exp(2i\pi f_n t). \tag{1.20}$$

Comme précédemment, cette opération est linéaire et invariante par translation temporelle, et le système qui transforme s en v s'appelle un *filtre*[7]. La fonction $H(f)$ (qui n'a été ici considérée que pour les fréquences f_n, mais peut très bien être définie pour toutes les fréquences f) s'appelle la *fonction de transfert* du filtre.

Quel est l'intérêt du filtrage ? Essentiellement, de modifier la composition fréquentielle du signal. Si par exemple le son est jugé trop « brillant», on appliquera un filtre pour lequel les $|H(f_n)|$ sont faibles (voire nuls) pour les hautes fréquences, et de l'ordre de 1 pour les basses fréquences. Un tel filtre est appelé filtre *passe-bas*. On peut aussi faire le contraire : atténuer les basses fréquences ; on a alors affaire à un filtre *passe-haut*. On peut enfin vouloir sélectionner des fréquences intermédiaires et atténuer les autres, auquel cas le filtre est dit *passe-bande*. Nous y reviendrons un peu plus loin.

Par exemple, le signal à peu près triangulaire de la figure 1.14 (courbe du bas) formé des quatre premiers harmoniques, a été filtré par chacun des trois filtres dont les fonctions de transfert sont représentées dans la colonne de gauche de la figure 1.29 : un passe-bas, un passe-bande et un passe-haut. Dans la colonne de droite sont représentés les signaux de sortie. Pour les deux premiers, ce filtrage a permis d'isoler respectivement les composantes de fréquence 1Hz et 2Hz. Le troisième filtre a isolé la somme des deux composantes de fréquence 3Hz et 4Hz.

Dans la figure 1.30, la même opération a été répétée sur le signal à peu près rectangulaire de la figure 1.15. Pouvez-vous l'interpréter ?

---

[7] On peut en fait montrer que toute opération linéaire et invariante peut s'écrire sous cette forme.

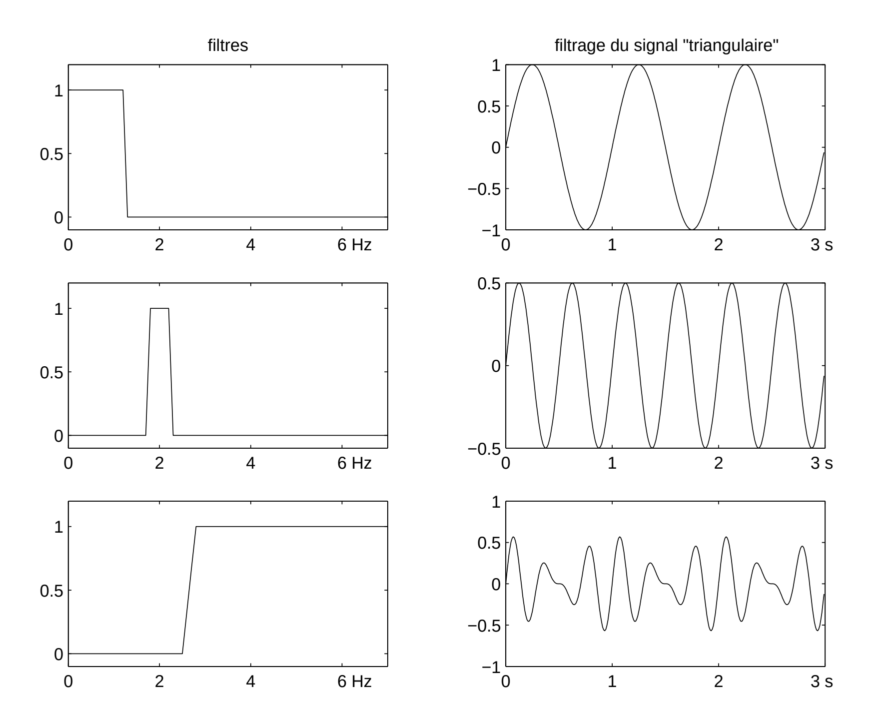


FIG. 1.29 – différents filtrages du signal « triangulaire».

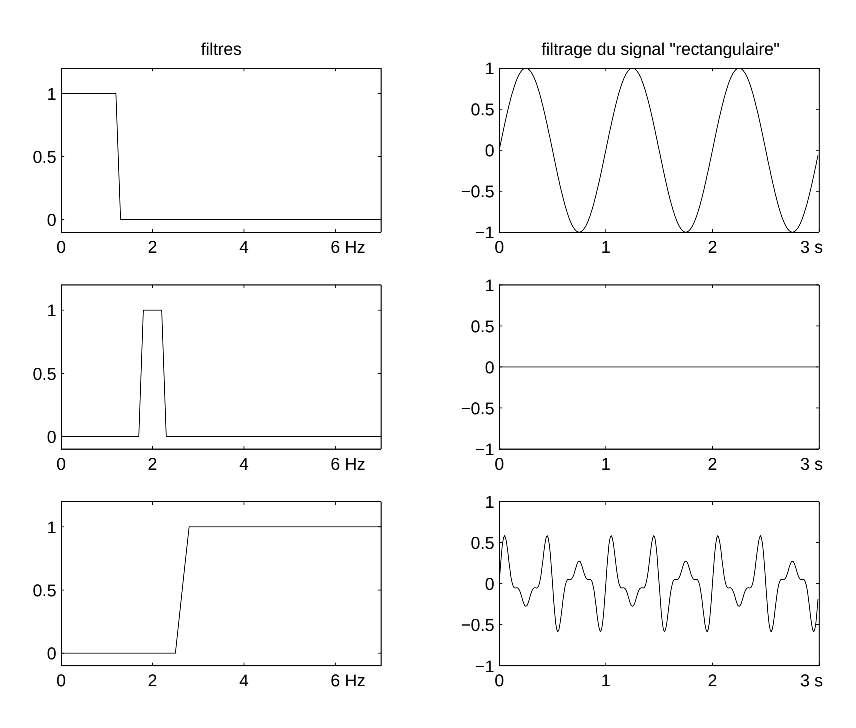


FIG. 1.30 – différents filtrages du signal « rectangulaire». Que se passe-t-il ?

Réponse impulsionnelle

Par ailleurs, on peut montrer que sous certaines conditions, $H(f)$ est la transformée de Fourier d'une fonction intégrable $h(t)$:

$$H(f) = \widehat{h}(f). \tag{1.21}$$

La fonction $h(t)$ est appelée la *réponse impulsionnelle* du filtre : c'est le signal que l'on trouve en sortie du filtre lorsque l'entrée est *l'impulsion de Dirac* en 0. Cette dernière, notée $\delta(t)$, est un signal infiniment bref (qui n'est pas une fonction !), dont la « masse» est concentrée en 0, et qui vérifie la relation

$$\int_{-\infty}^{+\infty} \varphi(t)\delta(t)\, dt = \varphi(0) \tag{1.22}$$

pour toute fonction φ continue et nulle en dehors d'un intervalle borné (voir aussi § 5.1). On montre alors que v dans (1.20) est le *produit de convolution* de h par s :

$$v(t) = (h * s)(t) := \int_{-\infty}^{+\infty} h(t-u)s(u)\, du. \tag{1.23}$$

1.5.2 Cas du spectre continu

Venons-en maintenant au filtrage de sons à spectre continu, c'est-à-dire rappelons-le de la forme

$$s(t) = \int_{-\infty}^{+\infty} \widehat{s}(f) \exp(2i\pi ft)\, df$$

où $\widehat{s}$ est la transformée de Fourier de s (cf. équation (1.15)). Le module $|\widehat{s}(f)|$ donne l'amplitude de la composante fréquentielle $\exp(2i\pi ft)$, et l'argument de $\widehat{s}(f)$ donne la phase à l'origine de cette composante.

Comme pour les cas étudiés précédemment, le filtrage de s par un filtre dont la fonction de transfert est $H(f)$ donnera un signal $v(t)$ qui aura pour transformée de Fourier $H(f)\widehat{s}(f)$:

$$\widehat{v}(f) = H(f)\widehat{s}(f). \tag{1.24}$$

La fonction $H(f)$ est la transformée de Fourier d'une fonction $h(t)$, toujours appelée réponse impulsionnelle, et l'on a encore le produit de convolution

$$v = h * s. \tag{1.25}$$

La théorie mathématique sous-jacente a été développée au XIXème siècle et au début du XXème siècle, en particulier par Laurent Schwartz (1915-2002), « l'inventeur» des distributions.

Application : le son que reçoit l'auditeur dans une salle de concert peut être vu comme le résultat du filtrage du son provenant de l'orchestre par le filtre que constitue la salle elle-même, dont on peut avoir une idée de la réponse impulsionnelle en frappant dans ses mains ou en émettant un son très bref. Selon vous, que percevra l'auditeur si la fonction de transfert de la salle ressemble au graphique de la figure 1.31 ?

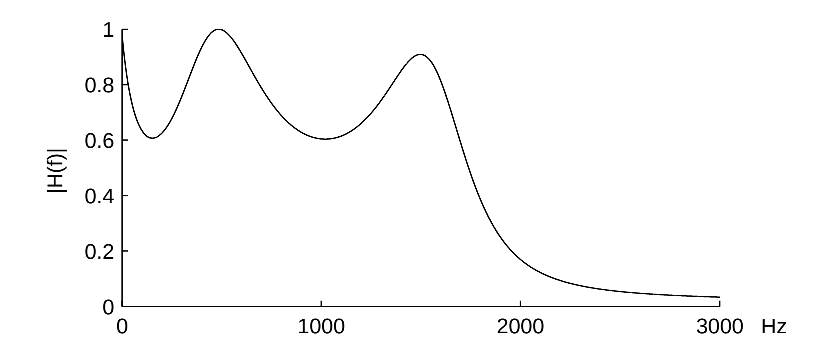


FIG. 1.31 – fonction de transfert d'une salle.

1.5.3 Filtres idéaux passe-bas, passe-bande et passe-tout

Terminons ce chapitre par la description de trois modèles de filtres fondamentaux, que nous aurons l'occasion d'utiliser à plusieurs reprises.

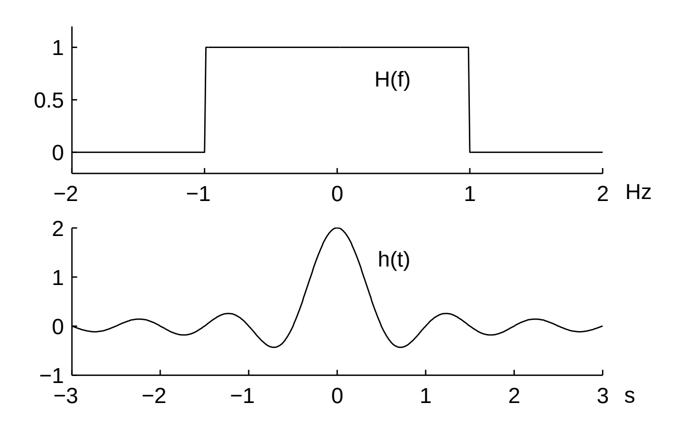


FIG. 1.32 – fonction de transfert et réponse impulsionnelle du filtre passe-bas idéal de fréquence de coupure $B = 1$.

Le *filtre passe-bas idéal*[8] de *fréquence de coupure* $B > 0$ (fig. 1.32) est donné par sa fonction de transfert[9] :

$$H(f) = \begin{cases} 1 \text{ si } |f| < B \\ 0 \text{ sinon.} \end{cases} \tag{1.26}$$

Ce filtre laisse donc passer sans modification les fréquences $|f| < B$ et bloque les fréquences $|f| > B$ (on ne peut rien dire du cas limite $|f| = B$).

[8]dit « idéal» car non réalisable physiquement. On peut simplement l'approcher d'aussi près que l'on veut par un système électronique, à condition d'accepter un certain retard sur la sortie.

[9]La valeur de H aux deux extrémités B et $-B$ n'a en principe aucune importance, car l'intégrale ne « voit» pas les valeurs ponctuelles isolées, sauf s'il y a des masses de Dirac à cet endroit, mais ceci est une autre histoire...

En se rappelant que $H(f) = \widehat{h}(f)$, la réponse impulsionnelle h s'obtient en utilisant (1.14) :

$$h(t) = \int_{-B}^{B} 1 \exp(2i\pi f t)\, df = \frac{\sin(2\pi B t)}{\pi t}.$$

Ainsi, on a

$$h(t) = 2B\,\mathrm{sinc}(2Bt),$$

où la fonction *sinus cardinal* (« cardinal» car s'annulant sur les *entiers* $\neq 0$) est donnée par

$$\mathrm{sinc}(t) = \frac{\sin(\pi t)}{\pi t}.$$

Le *filtre passe-bande idéal* de *fréquences de coupure* $f_0 - B > 0$ et $f_0 + B$ (fig. 1.33) est donné par sa fonction de transfert :

$$H(f) = \begin{cases} 1 \text{ si } |f \pm f_0| < B \\ 0 \text{ sinon.} \end{cases} \tag{1.27}$$

Ce filtre laisse donc passer sans modification les fréquences intermédiaires $|f \pm f_0| < B$ et bloque les autres fréquences.

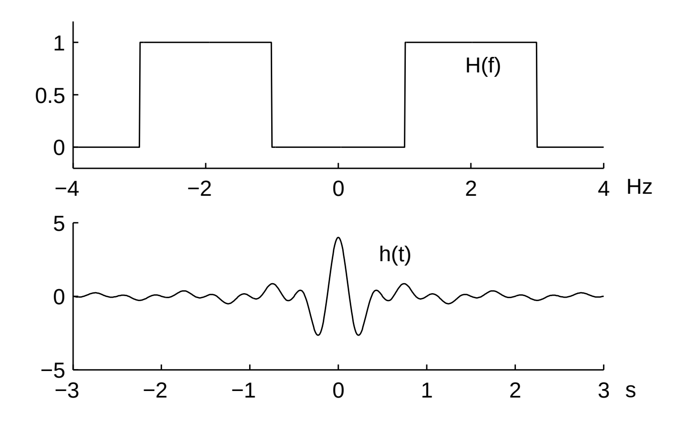


FIG. 1.33 – fonction de transfert et réponse impulsionnelle du filtre passe-bande idéal de fréquences de coupure 1 et 3.

Le calcul de sa réponse impulsionnelle donne

$$h(t) = 4B\,\mathrm{sinc}(2Bt)\cos(2\pi f_0 t).$$

On remarquera que celle-ci est une fréquence $\cos(2\pi f_0 t)$ (dite *porteuse* en communications radio) modulée en amplitude par la réponse impulsionnelle du filtre passe-bas idéal.

Enfin, le *filtre passe-tout*, au nom étrange, laisse tout passer ! Sa fonction de transfert est de module 1, et donc de la forme

$$H(f) = \exp(-i\theta(f)), \quad \theta(f) \in \mathbb{R}.$$

Un son pur $\exp(2i\pi f t)$ passant dans ce filtre se transforme en $\exp(2i\pi f t - i\theta(f))$: il a donc subi un déphasage d'angle $\theta(f)$, variable selon la fréquence, mais aucune modification d'amplitude. Ce genre de filtre s'utilise par exemple pour simuler la réverbération (cf. chapitre 6).

1.6 Sujets d'étude

Chaque chapitre est accompagné de sujets d'étude qui permettent d'approfondir certains points effleurés dans le cours. La difficulté et la quantité de travail demandée sont indiquées par des astérisques : (*) pour facile, (**) pour moyen et (***) pour difficile.

1.6.1 Réflexion normale sur un mur (*)

Dans le demi-espace $x \geq 0$, de coordonnées x, y, z, une onde plane incidente $p_i(x,t) = \sin(kx + 2\pi ft)$ est réfléchie par un mur. Ce mur, d'équation $x = 0$, est supposé parfaitement rigide. La vitesse de l'air est donc nulle en $x = 0$. L'onde plane incidente produit une pression acoustique

$$p_a(x,t) = p_i(x,t) + p_r(x,t)$$

où $p_r(x,t) = \beta \sin(kx - 2\pi ft)$ est l'onde réfléchie.
1) En utilisant l'équation d'Euler, montrez que $\partial_x p_a(0,t) = 0$ pour tout t.
2) Calculez la valeur de β et montrez que $p_a(x,t) = 2\cos(kx)\sin(2\pi ft)$. Quelle est la nature de cette onde plane ?

1.6.2 Filtrage en peigne par un micro situé à proximité d'un mur (**)

Lors de l'enregistrement d'un son pur de fréquence f, engendrant une onde acoustique harmonique, un microphone est placé suffisamment près d'un mur pour que l'on puisse considérer que l'onde produite y est plane. On suppose ainsi que les conditions du sujet d'étude 1.6.1 sont satisfaites.
1) Quelle sera, en un point donné, l'intensité sonore sur une période $T = 1/f$?
2) Le microphone, placé à une distance d du mur, est supposé ne réagir qu'aux variations de pression. Pour quelles valeurs de la fréquence l'amplitude $\alpha(f)$ du signal perçu par le microphone sera-t-elle maximale ? minimale ? quel est l'effet de la distance d sur ces valeurs ? Représentez graphiquement $10\log(\alpha^2(t))$.
3) Même question si l'on suppose que le microphone ne réagit qu'aux variations de vitesse (cf. [6] pour en savoir plus sur les micros).

1.6.3 Addition des intensités (***)

Un auditeur est placé suffisamment loin de l'orchestre pour que l'on puisse considérer que les 10 violons de cet orchestre produisent chacun à son niveau une onde de pression plane progressive

$$p_i(x,t) = u_i(x - ct), \quad i = 1, 2 \ldots, 10.$$

1) En utilisant l'équation d'Euler et l'équation d'état, montrez que la vitesse de l'air correspondante est $v_i(x,t) = p_i(x,t)/c\rho_0$.

L'auditeur est placé en un point x fixé. Chacun des violons joue la même note, de fondamental f, avec la même force, de sorte que

$$p_i(x,t) = s(t - \varphi_i),$$

où les φ_i reflètent les différences de phase qu'il peut y avoir entre les sons au point x, la fonction s étant T-périodique avec $T = 1/f$.
2) Calculez d'abord, en fonction de s, l'intensité sonore I_1 produite en ce point par un seul violon sur une période T.
3) par quel facteur devrait-on multiplier l'amplitude de ce violon pour obtenir une augmentation de l'intensité de 10dB ? De 20dB ? (réponse : par $\sqrt{10} \simeq 3.16$; par 10).
4) Les 10 violons jouent maintenant tous ensemble. Au point x, la pression acoustique totale et la vitesse associée sont donc

$$p_a(x,t) = \sum_{i=1}^{10} p_i(x,t), \quad v(x,t) = \sum_{i=1}^{10} v_i(x,t).$$

Montrez que l'intensité totale I_{10} au point x vaut

$$I_{10} = \frac{1}{c\rho_0 T} \int_0^T \left(\sum_{i=1}^{10} s(t - \varphi_i) \right)^2 dt.$$

En utilisant l'inégalité de Cauchy-Schwarz

$$\int_0^T g(t)h(t)dt \leq \left(\int_0^T g^2(t)dt \right)^{1/2} \left(\int_0^T h^2(t)dt \right)^{1/2},$$

déduisez-en que l'on a $0 \leq I_{10} \leq 100 I_1$, et que les niveaux en décibel associés vérifient $L_{I_{10}} \leq L_{I_1} + 20\text{dB}$.
5) A quelle(s) situation(s) correspond le cas $L_{I_{10}} = L_{I_1} + 10\text{dB}$ mentionné dans le cours ?

1.6.4 Intensité d'une onde stationnaire (**)

Soit une onde de pression stationnaire, de la forme

$$p_a(x,t) = a \sin(k(x - x_0)) \cos(2\pi f(t - t_0)).$$

En utilisant l'équation d'Euler pour déterminer la vitesse v, montrez que l'intensité sur une période $T = 1/f$ est nulle. Interprétation : une onde stationnaire ne transporte pas d'énergie (celle-ci ne fait qu'osciller sur place).

1.6.5 Son d'une sirène (*)

La sirène fut inventée par l'ingénieur français Cagniard de La Tour (1777-1859). Pour un son de la forme $s(t) = \sin(2\pi F(t))$, on appelle *fréquence instantanée* la fonction $f(t) = F'(t)$. Déterminez l'expression $s(t)$ d'une sirène dont la fréquence instantanée varie de manière sinusoïdale entre deux fréquences $f_0 - \beta$ et $f_0 + \beta$.

1.7 Travaux pratiques sur ordinateur

Cette première série de travaux pratiques est consacrée à la fabrication, l'écoute et l'analyse de sons en utilisant le logiciel MATLAB.

Convention typographique : les objets mathématiques sont écrits en *italique* (par exemple, le son $s(t)$). Les objets MATLAB sont écrits en style `typewriter` (par exemple, le troisième élément `s(3)`).

1.7.1 Premier son, les vecteurs

Sur ordinateur, un son $s(t)$ est représenté par ses valeurs (dites *discrètes*) $s_n = s(t_n)$ avec $t_n = n\tau$ et $\tau = 1/F_e$. Deux instants consécutifs t_n et t_{n+1} sont ainsi séparés par un petit intervalle de temps τ appelé *période d'échantillonnage,* et le nombre d'échantillons par seconde est égal à F_e, que l'on appelle *fréquence d'échantillonnage* (ces notions sont abordées en détail au chapitre 5).

Dans MATLAB, les valeurs s_n peuvent être rangées dans un vecteur `s`, et l'accès à l'élément numéro n se fait en écrivant `s(n)`. Exemple de création d'un vecteur et d'accès à l'un de ses éléments : après avoir lancé MATLAB, écrivez les lignes suivantes (le signe ≫ est le « prompt» qui apparaît lorsqu'on est dans MATLAB), en appuyant sur la touche « entrée» à la fin de chaque ligne :

```
≫ s = [1,-0.5,2,3] ;
≫ s
≫ s(3)
```

Attention : la virgule sert de séparateur entre les nombres, et le point est la notation anglo-saxonne pour la virgule des nombres décimaux. Notez que le résultat de l'opération *ne s'affiche pas* ou *s'affiche* selon que la ligne se termine par un point-virgule (cas de la première ligne) ou non (lignes 2 et 3).

Pour générer un vecteur $x = [a, a + h, a + 2h, \ldots, a + nh]$, dont les éléments sont équidistants, c'est très simple : il suffit d'utiliser la commande `x = a :h :(a+n*h) ;` après avoir affecté des valeurs aux variables `a`, `h` et `n`.

Premier son (son pur) : interprétez et exécutez les lignes de commande suivantes :

```
≫ Fe = 22050 ;
≫ f = 440 ;
≫ T = 1 ;
≫ dt = 1/Fe ;
≫ t = 0 :dt :T ;
≫ s = sin(2*pi*f*t) ;
≫ sound(s,Fe) ;
```

Un outil indispensable : l'aide en ligne. Pour avoir une description sommaire d'une fonction MATLAB, il suffit d'exécuter la commande `help` suivie du nom de la fonction, par exemple `help sin` ou `help sound`. Et si vous voulez de l'aide sur l'aide, tapez `help help` ! Une particularité de la plupart des fonctions MATLAB est de retourner un vecteur si l'argument est lui-même un vecteur. Par exemple, le vecteur `s` ci-dessus est composé des valeurs $s(t_n) = \sin(2\pi f t_n)$ sur l'intervalle $[0, T]$ qui a été spécifié.

1.7.2 Faisons varier les paramètres : le fichier de commandes

Si l'on veut faire varier la fréquence par exemple, c'est peu pratique de tout récrire. Il vaut mieux utiliser un fichier dans lequel on écrit toutes les commandes. Ouvrez un fichier

que vous nommerez `test1.m`, et écrivez dans ce fichier la liste des commandes précédentes (sans le ≫!). Pour exécuter l'*ensemble de toutes les commandes* contenues dans ce fichier, il suffit alors de taper la commande `test1` sous le prompt de MATLAB, après avoir bien sûr enregistré le fichier.

Modifiez maintenant à votre gré les valeurs de `Fe`, `T` et `f` dans le fichier `test1.m`, et interprétez ce que vous entendez à l'exécution du fichier.

1.7.3 Créons des sons plus complexes : utilisation des fonctions

Nous voulons maintenant créer un son plus complexe, comportant plusieurs fréquences. On peut poursuivre l'écriture dans le fichier de commande précédent, mais il est plus commode d'utiliser une *fonction* si l'on veut faire des essais variés.

À la différence des fichiers de commande, les fonctions renvoient le(s) résultat(s) d'un calcul utilisant un ou plusieurs arguments. Ces fonctions sont également écrites dans un fichier ayant l'extension `.m`, mais la première ligne, appelée *en-tête,* est obligatoirement de la forme

```
function [y1, y2, ..., yp] = fonct(x1, x2, ..., xq)
```

où `fonct.m` est le nom du fichier.

Télé-chargez (en cliquant sur le bouton droit de la souris), ouvrez et lisez le fichier `synthad.m` qui se trouve sur ma page web `www.gmm.insa-tlse.fr/~guillaum` à la rubrique `TP son et musique`. Les lignes commençant par % sont des commentaires.

Exécutez par exemple les commandes suivantes (les autres arguments conservant leurs valeurs antérieures) :

```
≫ a = 1 ; p = 0 ;
≫ s = synthad(a,f,p,T,Fe) ;
≫ sound(s,Fe) ;
```

Grâce à cette fonction, nous allons pouvoir générer des sons plus complexes : si l'on passe comme arguments à `synthad` les vecteurs $a = [\alpha_1, \alpha_2, \ldots, \alpha_m]$, $f = [f_1, f_2, \ldots, f_m]$, $p = [\theta_1, \theta_2, \ldots, \theta_m]$ et les nombres T, F_e, la fonction retournera dans le vecteur `s` les échantillons $s(t_n)$ du son

$$s(t) = \sum_{n=1}^{m} \alpha_n \sin(2\pi f_n t + \theta_n), \quad 0 \leq t \leq T.$$

Quelques précisions techniques avant d'aller plus loin :

Opérations sur les vecteurs. La transposition d'un vecteur ligne en vecteur colonne ou réciproquement se fait en ajoutant une apostrophe : `x'`. L'addition ou la soustraction de vecteurs de même taille se fait avec les opérateurs + et -. Plus curieux : l'addition `a+x` d'un nombre `a` et d'un vecteur `x` ajoute `a` à chaque composante de `x`. La multiplication ou la division d'un vecteur `x` par un nombre `a` se fait en écrivant `a*x` et `x/a`. MATLAB dispose d'outils pratiques pour effectuer des opérations sur les vecteurs sans effectuer de boucles : ce sont les opérateurs `.*` et `./` qui opèrent *terme à terme.* Pour élever tous les termes d'un vecteur `x` à la puissance `m`, on écrit `x.^m`. Pour illustrer cela, tapez les commandes suivantes :

```
≫ x = [1,2,3] ; y = [2,2,3] ;
≫ x
≫ x'
```

```
≫ x+y
≫ x'+y
≫ x+0.1
≫ 3*x
≫ x/2
≫ x.*y
≫ x./y
≫ x.^2
≫ x.^y
```

La ligne `x'+y` a provoqué une erreur ! On ne peut additionner deux vecteurs que s'ils ont même forme : ligne ou colonne. À vous maintenant de jouer sur les valeurs des amplitudes et des fréquences que vous passerez à la fonction `synthad`, et de comparer les sons obtenus. Pour le cas harmonique, vous pouvez en particulier jouer sur la décroissance plus ou moins rapide des coefficients α_n, ou sur la présence ou non des harmoniques pairs. Pour le cas des partiels, essayez les valeurs du cours données pour une cloche.

Intermède bruits et sirènes

Sur le modèle de `synthad.m`, créez une fonction `bruit.m` ayant pour en-tête `function s = bruit(T,Fe)`, et fabriquant un bruit, soit sur l'un des modèles décrits en cours, soit tout simplement en utilisant la fonction `randn` de MATLAB (pensez au `help`). Créez également une fonction `sirene.m`, d'en-tête `function s = sirene(f,f1,beta,T,Fe)`, et renvoyant les échantillons du son

$$s(t) = \sin(2\pi f t + \beta \sin(2\pi f_1 t)/f_1).$$

1.7.4 Analyse

Vous allez maintenant analyser un son de votre choix, que vous aurez crée ou télé-chargé parmi les fichiers son (extension `.wav`) que vous trouverez sur ma page web. Pour lire par exemple le fichier `flute.wav`, utilisez la commande `[s,Fe] = wavread('flute.wav')`. À noter que le vecteur `s` se présente sous forme d'une colonne.

Ci-dessous un exemple de son composé de trois notes consécutives comportant chacune trois harmoniques (nous ne ferons désormais plus apparaître le prompt ≫) :

```
a = [100,10,1] ;
f = [440,880,1320] ;
p = [0,0,0]+pi/2 ;
T = 1 ; Fe = 11025 ;
s1 = synthad(a,f,p,T,Fe) ;
s2 = synthad(a,1.5*f,p,T,Fe) ;
s3 = synthad(a,2*f,p,T,Fe) ;
s = [s1,s2,s3] ;
soundsc(s,Fe) ;
```

L'avant dernière ligne donne un vecteur `s` résultant de la mise bout à bout des trois vecteurs `s1,s2,s3`. Nous avons ici utilisé la fonction `soundsc` plutôt que la fonction `sound` qui ne marche correctement que si toutes les valeurs sont comprises dans l'intervalle $[-1, 1]$. La

fonction `soundsc` (`sc` pour 'scale', c'est-à-dire mise à l'échelle) multiplie automatiquement le son par un nombre approprié pour que toutes les valeurs soient dans cet intervalle.

Analyse temporelle

C'est la représentation temporelle du signal, qui se fait par la commande `plot(x)`. Vous pouvez zoomer en appuyant sur l'icône ⊕ et en délimitant la zone à agrandir avec le bouton gauche de la souris. Zoom arrière en cliquant sur le bouton droit de la souris.

Analyse fréquentielle

Pour effectuer l'analyse fréquentielle d'un son $s(t)$ de durée T, sur la bande de fréquence $[0, F_{\max}]$ où l'on impose $F_{\max} < F_e/2$ (l'explication viendra au chapitre 5), exécutez les commandes suivantes, où l'on a pris ici $F_{\max} = 4000$:

```
z = fft(s) ;
T = (length(s)-1)/Fe ;
fr = 0 :1/T :4000 ;
nf = length(fr) ;
plot(fr,abs(z(1 :nf))) ;
```

Si par exemple vous avez pris pour `s` un son périodique ayant pour harmoniques 440Hz, 880Hz et 1320Hz, vous devez voir apparaître trois raies d'abscisses 440, 880 et 1320, de hauteur proportionnelle aux poids (les α_n) de chacun de ces harmoniques.

Analyse temps-fréquence

L'analyse temps-fréquence se fait en utilisant la fonction `specgram` de MATLAB :

```
specgram(s,512,Fe) ;
```

Le nombre 512 indique la largeur de la fenêtre d'analyse (cf. § 1.4.2), et la vitesse de calcul de la FFT est optimale si ce nombre est une puissance de 2. Essayez différentes valeurs et interprétez les spectrogrammes obtenus.

1.7.5 Filtrage

Nous allons filtrer le son `steinwayE.wav` à télé-charger sur ma page et à lire dans MATLAB en utilisant la commande `[s,Fe] = wavread('steinwayE.wav')`. Nous admettrons que les lignes suivantes réalisent un filtrage *passe-bas* de fréquence de coupure $W = W_n * F_e/2$ où $W_n \in]0, 1[$:

```
b1 = fir1(100,Wn) ;
s1 = filter(b1,1,s) ;
```

Faites varier `Wn` de telle sorte que W varie entre 100Hz et 1000Hz et écoutez le résultat.

De même, nous admettrons que les lignes suivantes réalisent un filtrage *passe-haut* de même fréquence de coupure $W = W_n * F_e/2$:

```
b2 = fir1(100,Wn,'high') ;
s2 = filter(b2,1,s) ;
```

Faites également varier W dans l'intervalle précédent et écoutez le résultat.

Enfin, effectuez un filtrage passe-bande de votre choix en consultant l'aide en ligne de la fonction `fir1`.

Chapitre 2

Les instruments

Après avoir étudié la propagation et la nature des sons musicaux, allons maintenant à la source de ces sons : les instruments de musique. La compréhension du mode de production des sons musicaux a toujours été une préoccupation majeure des compositeurs, musiciens et scientifiques. Jean-Philippe Rameau affirmait par exemple que « la simple résonance du Corps Sonore donne la loi à toute musique théorique et pratique». De nombreux mathématiciens et physiciens y ont apporté leur contribution [3]. Cependant, nous sommes bien loin de disposer de modèles complets permettant de décrire de manière exacte les instruments de musique. Leur fonctionnement peut être extrêmement complexe et faire appel à des théories très sophistiquées, comme celle de la turbulence pour décrire les oscillations du filet d'air produit au niveau de l'embouchure d'une flûte ou d'un tuyau d'orgue [20, 4].

De manière simplifiée, un instrument de musique comporte deux parties essentielles : *l'excitateur* (la source des vibrations) et le *résonateur*. Une corde seule ne produit qu'un son à peine audible. Il faut lui associer un résonateur pour mettre l'air en mouvement, transformer l'énergie mécanique de la vibration en énergie acoustique. Il est possible que les hommes préhistoriques se soient rendus compte de ce phénomène en se servant de leur bouche pour tendre un arc : une peinture rupestre de la grotte des Trois-Frères (Ariège), vielle de 10000 à 15000 ans, représente un sorcier tenant la partie supérieure de l'arc entre les dents, la partie inférieure par la main gauche, et jouant sur la corde de la main droite. Dans la majorité des instruments à cordes, le résonateur est une plaque en épicéa, appelée *table d'harmonie*, une pièce de bois (hêtre ou érable) appelée *chevalet* servant d'intermédiaire entre la corde et la table. Dans les vents, l'excitateur peut être une anche en roseau (clarinette, saxophone, hautbois...), les lèvres de l'instrumentiste (cor, trompette, trombone...), un jet d'air (flûte, orgue...), et le résonateur est la colonne d'air délimitée par l'instrument, parfois également le tuyau lui-même, selon le matériau dont il est fait.

L'objectif de ce chapitre est d'étudier les modèles simples de vibration des corps sonores, qui sont à la base du fonctionnement de tous les instruments de musique. Ces vibrations peuvent se classer en deux catégories : les vibrations *libres* et les vibrations *entretenues.* Dans la première catégorie, on trouve les instruments à percussion, le clavecin, le piano, la guitare, mais aussi le violon lorsque l'on joue un *pizzicato* ! Les vibrations sont dites libres car après une action brève (percussion, pincement), le corps n'est plus soumis à aucune action et continue de vibrer librement. Dans la seconde catégorie, on trouve les instruments à vent (bois, cuivres, orgue) et les instruments à cordes frottées (violon, violoncelle, contrebasse). Le

son y est entretenu par l'action du souffle pour les vents ou de l'archet pour les cordes. Nous étudierons la vibration libre des cordes, des lames et des membranes, la vibration entretenue dans un tube, et nous verrons en quoi les différents types de fonctionnement induisent les propriétés caractéristiques des instruments : leur timbre.

2.1 Les cordes

Examinons une corde de longueur L et de masse linéique μ (masse par unité de longueur), fixée à ses deux extrémités et soumise à une tension[1] T. Trois types de vibrations sont possibles : transversales, longitudinales et en torsion. Nous n'étudions que le premier type, et supposons que la vibration a lieu dans un plan xOy. Dans ce plan, la corde a pour extrémités les points $(0,0)$ et $(0,L)$, et la position de la corde à l'instant t est donnée par l'équation $y = u(x,t)$. Les conditions aux limites imposent $u(0,t) = u(L,t) = 0$ pour tout t.

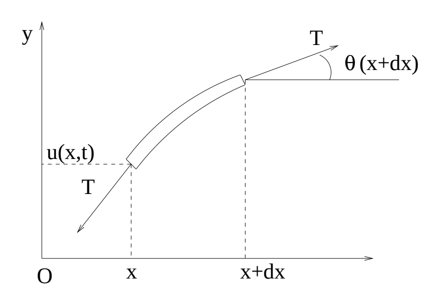


Fig. 2.1 – segment de corde en plein effort.

Pour obtenir l'équation régissant le mouvement de la corde, on considère, à un instant donné t, les forces qui agissent sur un petit segment de corde situé entre les abscisses x et $x + dx$ (cf. fig. 2.1). On note $\theta(x)$ l'angle que fait la corde avec l'axe Ox. Au point $x + dx$, la composante verticale de la force due à la tension est

$$F(x + dx) = T \sin\theta(x + dx) \simeq T\,\mathrm{tg}\theta(x + dx) = T\partial_x u(x + dx, t),$$

l'approximation étant valable si $\theta(x)$ est proche de 0. Au point x, on trouve de même

$$F(x) \simeq -T\partial_x u(x,t).$$

La relation fondamentale de la dynamique $F = m\gamma$ donne pour l'accélération verticale

$$T\partial_x u(x + dx, t) - T\partial_x u(x,t) \simeq \mu dx \partial_{t^2} u(x,t).$$

En divisant par dx et en faisant tendre dx vers 0, nous retrouvons l'équation des ondes (1.6)

$$\partial_{t^2} u(x,t) = c^2 \partial_{x^2} u(x,t)$$

[1]En notant S et ΔL la section et l'allongement de la corde, on a $T = SE\Delta L/L$ où E est le module de Young du matériau de la corde (de l'ordre de $2\,10^{11}$Pa pour l'acier, avec $1\,\mathrm{Pa} = 1\,\mathrm{Nm}^{-2}$).

où

$$c = \sqrt{T/\mu} \tag{2.1}$$

est maintenant (cf. l'analyse faite au § 1.1.2) la vitesse de propagation d'une onde transversale progressive parcourant la corde (à ne pas confondre avec la vitesse d'un point de la corde). Comme nous l'avons fait dans le chapitre 1, on peut commencer par chercher les solutions harmoniques

$$u(x,t) = \varphi(x)\exp(2i\pi ft)$$

où $\varphi(x)$ est solution de l'équation de Helmholtz (avec $k = 2\pi f/c$)

$$\varphi''(x) + k^2\varphi(x) = 0. \tag{2.2}$$

Les solutions de cette équation sont de la forme

$$\varphi(x) = \alpha\exp(ikx) + \beta\exp(-ikx).$$

La prise en compte des *conditions aux limites* $\varphi(0) = \varphi(L) = 0$ donne le système linéaire homogène suivant :

$$\begin{cases} \alpha + \beta = 0, \\ \alpha\exp(ikL) + \beta\exp(-ikL) = 0, \end{cases}$$

qui ne peut avoir de solution non nulle que si le déterminant $\exp(-ikL) - \exp(ikL) = -2i\sin(kL)$ est nul, c'est-à-dire si[2]

$$\begin{aligned} k = k_n &:= \frac{n\pi}{L}, \quad n \in \mathbb{Z}^*, \\ f = f_n &:= \frac{nc}{2L}. \end{aligned} \tag{2.3}$$

C'est la loi de Taylor pour la corde vibrante. Si tel est le cas, on a alors $\alpha = -\beta$ et $\varphi(x)$ est proportionnelle à $\sin(k_n x)$. Nous obtenons ainsi toute une famille d'ondes stationnaires

$$u(x,t) = a\sin(k_n x)\exp(2i\pi f_n t).$$

La première fréquence est le fondamental f_1, et les autres fréquences f_n sont des multiples *entiers* de f_1 : ce sont donc des harmoniques. Ces fréquences de résonance sont appelées les *fréquences propres* du système, et les ondes harmoniques correspondantes les *modes propres*. Les trois premiers modes propres sont représentés sur la figure 2.2. La relation (2.3) nous dit en particulier que la fréquence est inversement proportionnelle à la longueur de la corde, ce qui avait déjà été observé par Pythagore (500 ans A.C.) qui partit de cette constatation pour construire la gamme diatonique à partir d'un cycle de quintes (cf. chapitre 3).

On peut alors montrer que toute solution de l'équation des ondes pour la corde fixée à ses deux extrémités s'obtient par addition des modes propres (cf. aussi § 2.6.1 pour une autre technique de résolution), et s'écrit donc sous la forme générale

$$u(x,t) = \sum_{n=-\infty}^{+\infty} c_n \sin(k_n x)\exp(2i\pi f_n t). \tag{2.4}$$

[2]Le cas $n = 0$ est écarté car donnant une solution nulle $u = 0$. À noter que les fréquences négatives se ramènent à des fréquences positives lorsque l'on repasse dans le domaine réel.

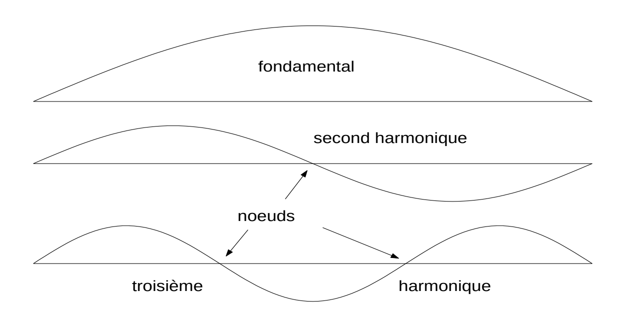


FIG. 2.2 – les trois premiers modes propres de la corde : Do, Do, Sol.

Les fréquences étant toutes des multiples entiers de f_1, il s'ensuit que le son produit est *périodique*, de période $1/f_1$. En fait, la réalité s'éloigne un peu de ce modèle simplifié, nous y reviendrons à la fin du chapitre dans la discussion sur le timbre.

Pour terminer, la détermination des coefficients c_n se fait en considérant les *conditions initiales* : la position $u_0(x)$ et la vitesse $v_0(x)$ de la corde à l'instant $t = 0$. La corde étant fixée aux extrémités, on a $u_0(0) = u_0(L) = 0$. En prolongeant $u_0(x)$ et $v_0(x)$ de façon *impaire* sur l'intervalle $[-L, L]$, on obtient deux fonctions $2L$-périodiques. Celles-ci se décomposent de manière unique en séries de sinus, qui coïncident avec $u_0(x)$ et $v_0(x)$ sur l'intervalle $[0, L]$:

$$u_0(x) = \sum_{n\geq 1} u_n \sin(n\pi x/L), \tag{2.5}$$

$$v_0(x) = \sum_{n\geq 1} v_n \sin(n\pi x/L). \tag{2.6}$$

La comparaison terme à terme de ces deux équations avec (2.4) et sa dérivée par rapport à t (la vitesse), en utilisant $k_n = n\pi/L = 2\pi f_n/c$ et $\sin(2\pi k_{-n}x) = -\sin(2\pi k_n x)$, donne pour chaque $n \geq 1$ le système

$$\begin{aligned} c_n - c_{-n} &= u_n, \\ cni\pi c_n/L + cni\pi c_{-n}/L &= v_n, \end{aligned}$$

de déterminant $2cni\pi/L \neq 0$, ce qui définit complètement les valeurs de c_{-n} et c_n.

2.1.1 Battements, accord et consonance

Nous décrivons ici l'utilisation des battements pour accorder un instrument, en prenant comme exemple le piano. Ce procédé s'applique naturellement à d'autres instruments comme le clavecin, l'harmonium, l'accordéon et l'orgue, mais aussi aux musiciens eux-même à l'intérieur d'une formation musicale. Nous négligeons ici l'inharmonicité qui sera décrite au § 2.5.1 et prise en compte au § 3.3.

Un piano comporte trois cordes par note, accordées à l'unisson, c'est-à-dire à la même fréquence (les graves n'ont qu'une ou deux cordes). D'après (2.1) et (2.3), le fondamental

d'une corde est donné par

$$f_1 = \frac{1}{2L}\sqrt{\frac{T}{\mu}}.$$

La longueur L et la masse linéique μ étant déterminées par le constructeur, l'accord se fait en jouant sur la tension T. À noter que la tension totale sur l'ensemble des cordes d'un piano de concert contemporain dépasse vingt tonnes. Le réglage de la tension de chaque corde se fait au moyen d'une cheville que l'on tourne pour obtenir la fréquence désirée. Pour l'accord, on utilise un diapason qui sert de référence (le La-3, qui peut varier entre 440Hz et 444Hz, cf. § 3.3.1), et l'on procède ensuite par comparaison, en utilisant le phénomène de *battement* que nous analysons maintenant.

Pour cela, prenons pour commencer deux cordes d'un unisson légèrement désaccordé, délivrant deux fondamentaux s_1 et s_2 de fréquences proches $f_1 < f_2$.

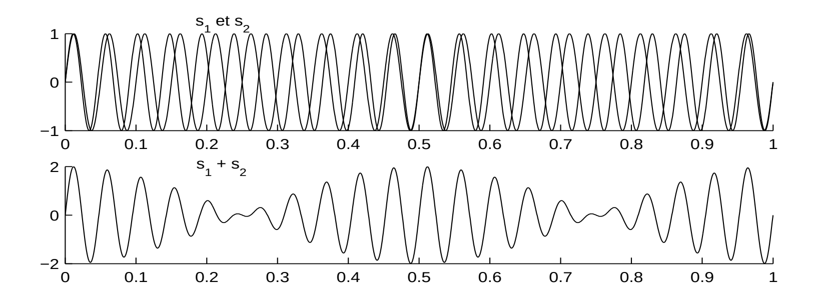


FIG. 2.3 – deux fréquences proches et leur superposition qui produit un *battement*.

Le graphique supérieur de la figure 2.3 représente les deux fondamentaux dans le cas

$$s_1(t) = \sin(20 \times 2\pi t), \quad s_2(t) = \sin(22 \times 2\pi t).$$

Le choix de ces fréquences a été fait pour la lisibilité des figures. Le graphique inférieur représente leur somme $s_1 + s_2$. Lorsque les deux sinusoïdes sont en phase, elles se renforcent ($t = 0$, 0.5 et 1). En opposition de phase aux instants $t = 0.25$ et $t = 0.75$, elles s'annulent mutuellement. C'est le phénomène de battement : le son s'amplifie et s'atténue périodiquement, à la fréquence $f_2 - f_1$ (ici 2Hz). Ce phénomène peut aussi se déduire de

$$\sin(2\pi f_1 t) + \sin(2\pi f_2 t) = 2\sin(2\pi\frac{f_1 + f_2}{2}t)\cos(2\pi\frac{f_1 - f_2}{2}t).$$

Le son résultant est un son de fréquence intermédiaire $(f_1 + f_2)/2$, *modulé en amplitude* par $\cos(2\pi t(f_1 - f_2)/2)$. Le maximum d'amplitude est atteint aux instants tels que $\cos(2\pi t(f_1 - f_2)/2) = \pm 1$, d'où une fréquence du battement égale à $2|f_1 - f_2|/2 = f_2 - f_1$.

Pour ajuster les deux cordes à la même fréquence, *il suffit donc de faire disparaître ce battement* (cf. aussi § 2.6.5). Ce n'est pas la *fréquence d'une corde individuelle* que mesure l'oreille de l'accordeur ou du musicien, mais *la fréquence (on dit aussi la vitesse) du battement* produit par deux cordes différentes. En effet, une différence d'un demi-hertz entre deux cordes jouées *consécutivement* est inaudible, même pour l'ouïe la plus fine, alors qu'un battement

d'un demi-Herz entre deux cordes jouées *simultanément* est parfaitement audible pour une oreille un tant soit peu entraînée. Le même procédé peut être utilisé pour accorder d'autres intervalles que l'unisson, à la différence que ce ne sont plus les battements entre fondamentaux qu'il faut utiliser, mais ceux qui se produisent entre les harmoniques, de rangs variables selon les notes à accorder. Examinons les trois cas illustrés pas la figure 2.4.

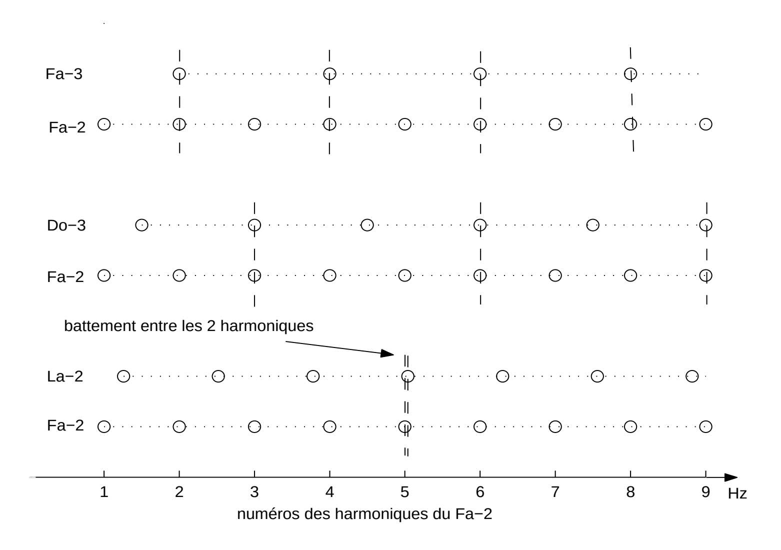


FIG. 2.4 – coïncidence d'harmoniques pour l'octave et la quinte, léger décalage pour la tierce majeure tempérée Fa-La au niveau des harmoniques respectifs 5 et 4.

- **Accord de l'octave.** Supposons accordé le Fa-2 à 174.6Hz (en tempérament égal). Le Fa-3 devrait être accordé à $2 \times 174.6 = 349.2$Hz. Supposons qu'il soit légèrement désaccordé, par exemple à 351.2Hz. Le second harmonique du Fa-2, qui a pour fréquence 349.2Hz, produira donc avec le fondamental du Fa-3 un battement de 2Hz. L'accord se fera alors en diminuant progressivement la fréquence du Fa-3 jusqu'à disparition de ce battement.
- **Accord de la quinte.** Considérons maintenant un Do-3 de fondamental théorique $3 \times 174.6/2$Hz. Son deuxième harmonique a pour fréquence 3×174.6Hz, ainsi que le troisième harmonique du Fa-2. L'accord de la quinte se fera donc en écoutant les battements entre ces deux harmoniques, jusqu'à leur disparition.
- **Accord de la tierce majeure.** Dans ce cas, l'accord ne se fait pas par suppression d'un battement, mais par *ajustement de sa vitesse.* Le Fa-2 produit un cinquième harmonique de fréquence 5×174.6Hz $= 873$Hz. Le La-2 produit lui un quatrième harmonique de fréquence 4×220Hz $= 880$Hz. Ces deux harmoniques produisent donc ensemble un battement de 7Hz perceptible par une oreille attentive. La vitesse de ce battement dépend bien sûr de la hauteur de la tierce. Par exemple, il sera deux fois plus rapide une octave plus haut. On remarque qu'en raccourcissant la tierce majeure (ici avec un Fa-2 de fréquence 880Hz$/5 = 176$Hz), on obtient une tierce sans battements, utilisée par exemple dans la gamme non tempérée de Zarlino [1, 7]. Nous reviendrons sur le tempérament au chapitre 3.

Si comme le fit Helmholtz au XIXème siècle, on définit le degré de *consonance* comme décrivant à quel point les harmoniques de deux notes jouées simultanément coïncident ou au contraire battent entre eux, l'octave est l'intervalle le plus consonant, suivi immédiatement par la quinte, d'où l'importance de cette derniere dans la construction des gammes.

2.2 Les lames

Quittons la section des cordes, et considérons maintenant le cas d'une tige ou d'une lame de section rectangulaire ou circulaire. C'est la source de vibration de nombreux instruments comme l'accordéon, le xylophone (« son de bois» en grec), le vibraphone (xylophone à lames métalliques (!), auquel on a rajouté des tubes qui servent de résonateur et un dispositif à clapets tournants pour faire varier l'amplitude de manière périodique), le celesta (tiges frappées par un marteau), le piano électro-acoustique Fender (*idem*), les boîtes à musique..., et dans les vents, l'anche elle-même !

Comme pour le cas des cordes, plusieurs types de vibrations sont possibles, et nous nous intéressons aux ondes transversales dans un plan xOy, l'axe central de la lame ayant pour extrémités les points $(0, 0)$ et $(0, L)$. La position de l'axe à l'instant t a pour équation $y = u(x, t)$. Le modèle mécanique est plus complexe que celui des cordes, et nous admettrons que le mouvement de la lame est régi par l'équation

$$\partial_{t^2} u(x,t) = -g^2 c_L^2 \partial_{x^4} u(x,t),$$

où g est le *rayon de giration* qui dépend de la forme de la section de la lame[3], $c_L = \sqrt{E/\rho}$ est la vitesse de propagation des ondes longitudinales dans la lame, E est le module de Young du matériau et ρ sa densité. Les solutions harmoniques sont toujours de la forme

$$u(x,t) = \varphi(x) \exp(2i\pi f t),$$

mais $\varphi(x)$ est maintenant solution d'une équation différentielle d'ordre 4 :

$$\varphi^{(4)}(x) = K^4 \varphi(x), \quad K \geq 0,\, K^4 = \left(\frac{2\pi f}{g c_L}\right)^2. \tag{2.7}$$

La solution générale de cette équation homogène s'écrit

$$\varphi(x) = \alpha \exp(Kx) + \beta \exp(-Kx) + \gamma \exp(iKx) + \delta \exp(-iKx), \tag{2.8}$$

les constantes étant déterminées par les conditions aux limites dont nous examinons les deux cas les plus courants pour les instruments de musique.

2.2.1 Lame posée aux deux extrémités

L'exemple type de lame posée aux deux extrémités est fourni par le xylophone. Dans le modèle simplifié où l'on suppose que la lame est posée exactement aux extrémités (ce qui

[3] $g = e/\sqrt{12}$ pour une lame d'épaisseur e, $g = r/2$ pour un cylindre de rayon r.

n'est pas très réaliste, mais passons), la fonction $\varphi(x)$ s'annule en 0 et L, ainsi que les dérivées secondes en l'absence d'efforts de flexion (cf. fig. 2.5). Ceci donne en 0 :

$$\begin{cases} \alpha+\beta+\gamma+\delta=0, \\ \alpha+\beta-\gamma-\delta=0, \end{cases}$$

d'où l'on déduit que $\alpha=-\beta$ et $\gamma=-\delta$, et

$$\varphi(x)/2=\alpha\,\mathrm{sh}(Kx)+i\gamma\sin(Kx).$$

Au point $x=L$, les deux autres conditions aux limites s'écrivent

$$\begin{cases} \alpha\,\mathrm{sh}(KL)+i\gamma\sin(KL)=0, \\ \alpha\,\mathrm{sh}(KL)-i\gamma\sin(KL)=0. \end{cases} \tag{2.9}$$

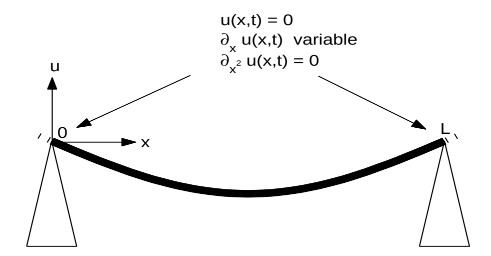


FIG. 2.5 – lame posée. Aux deux extrémités, la position est constante, la pente varie au cours du temps et la courbure est nulle.

On écarte le cas $K=0$, qui donnerait $u=0$. Le système (2.9) admet des solutions $(\alpha,\gamma)\neq(0,0)$ si et seulement si $\sin(KL)=0$, c'est-à-dire $K=n\pi/L$, avec $n\geq 1$ puisque $K>0$. Dans ces cas, on a $\alpha=0$ et, compte tenu de (2.7), les solutions harmoniques ou modes propres sont ainsi de la forme

$$u(x,t)=a\sin(K_nx)\exp(\pm 2i\pi f_nt) \tag{2.10}$$

avec $K_n=n\pi/L$ et $f_n=gc_LK_n^2\,/2\pi$, ou encore

$$f_n=n^2\frac{gc_L\pi}{2L^2}. \tag{2.11}$$

Remarque 2.1 *On constate que comme pour les cordes, les modes propres sont stationnaires et ont la même forme. Par contre, les fréquences propres f_n suivent une progression* quadratique

$$1,\ 4,\ 9,\ 16,\ldots,\ n^2,\ldots$$

qui s'oppose à la progression arithmétique *des fréquences propres de la corde. Remarque similaire concernant la longueur : les fréquences propres sont maintenant inversement proportionnelles au carré de la longueur. Dernière différence : on peut voir le mode (2.10) comme la*

superposition de deux ondes progressives $a[\exp(i(2\pi f_n t + K_n x)) - \exp(i(2\pi f_n t - K_n x))]/2i = u(x,t)$ *se propageant à la vitesse* $2\pi f_n/K_n = \sqrt{2\pi g c_L f_n}$, *qui maintenant* dépend de la fréquence. *On dit qu'il y a dispersion, car une onde progressive concentrée en espace ira en « s'étalant», les composantes hautes fréquences allant plus vite que les composantes basses fréquences.*

En superposant les parties réelles des solutions harmoniques, on obtient finalement les solutions physiques, qui s'écrivent sous la forme

$$u(x,t) = \sum_{n\geq 1} \alpha_n \sin(n\pi x/L)\cos(2\pi f_n t + \theta_n).$$

Toutes les fréquences étant des multiples entiers de la première fréquence, le résultat est un son périodique de période $1/f_1$, dans lequel certains harmoniques sont absents (octave, quinte au-dessus...), ce qui contribue à la sonorité particulière du xylophone. La résolution complète en fonction des données initiales s'effectue comme pour les cordes.

2.2.2 Lame encastrée à une extrémité

L'exemple type de lame encastrée à une extrémité et vibrant librement est fourni par la boîte à musique. La lame étant supposée encastrée en $x = 0$, la fonction $\varphi(x)$ (2.8) s'annule en 0 ainsi que sa dérivée. Au point L où la lame est libre, nous admettrons que ce sont les dérivées seconde et troisième qui s'annulent. Il est plus commode ici d'écrire $\varphi(x)$ sous la forme équivalente (mais avec des constantes différentes) :

$$\varphi(x) = \alpha\,\mathrm{ch}(Kx) + \beta\,\mathrm{sh}(Kx) + \gamma\cos(Kx) + \delta\sin(Kx).$$

Les deux conditions au point $x = 0$ donnent

$$\begin{cases} \alpha + \gamma = 0, \\ \beta + \delta = 0, \end{cases}$$

d'où

$$\varphi(x) = \alpha[\mathrm{ch}(Kx) - \cos(Kx)] + \beta[\mathrm{sh}(Kx) - \sin(Kx)].$$

Les deux conditions au point $x = L$ s'écrivent

$$\begin{cases} \alpha[\mathrm{ch}(KL) + \cos(KL)] + \beta[\mathrm{sh}(KL) + \sin(KL)] = 0, \\ \alpha[\mathrm{sh}(KL) - \sin(KL)] + \beta[\mathrm{ch}(KL) + \cos(KL)] = 0. \end{cases}$$

Ce système admet des solutions non nulles si et seulement si son déterminant est nul :

$$[\mathrm{ch}(KL) + \cos(KL)]^2 - \mathrm{sh}^2(KL) + \sin^2(KL) = 0,$$

qui après simplification donne

$$\frac{1}{\mathrm{ch}(KL)} + \cos(KL) = 0.$$

Notons λ_n les solutions positives (rangées par ordre croissant) de l'équation transcendante

$$\frac{1}{\mathrm{ch}(\lambda_n)} + \cos(\lambda_n) = 0. \tag{2.12}$$

Les solutions de (2.7) sont donc obtenues pour $K_n = \lambda_n/L$ et $f_n = gc_L K_n^2/2\pi$, soit

$$f_n = \lambda_n^2 \frac{gc_L}{2\pi L^2}, \tag{2.13}$$

et on peut vérifier qu'ici encore les modes propres sont stationnaires. Les quatre premiers sont représentés sur la figure 2.6.

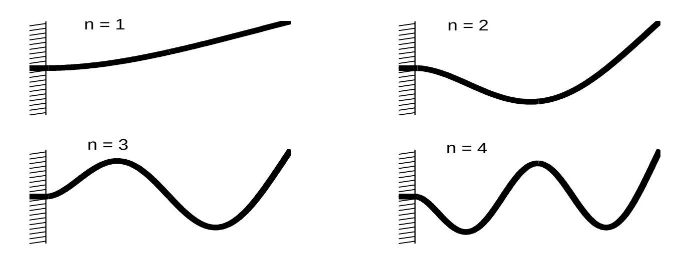


FIG. 2.6 – quatre premiers modes propres (amplifiés) d'une tige encastrée.

Contrairement au cas précédent (lame posée), les fréquences propres f_n ne sont plus des multiples *entiers* de la première fréquence f_1, ni d'ailleurs d'aucune autre fréquence f. Les fréquences f_n ne sont donc pas des harmoniques, mais des *partiels*. En résolvant numériquement (2.12), on trouve que les f_n sont proportionnels à la suite

$$1,\ 6.27,\ 17.55,\ 34.39, \ldots$$

On constate que la progression est (au début) plus rapide que pour la lame posée. Cependant, on peut déduire de (2.12) qu'asymptotiquement on a $\lambda_n \simeq (n - 1/2)\pi$, d'où pour n assez grand

$$f_n \simeq (n - 1/2)^2 \frac{gc_L\pi}{2L^2},$$

progression quadratique comparable à celle de la lame posée (2.11). L'accord des lames se fait en général en agissant sur l'épaisseur en retirant de la matière, par exemple au niveau du point d'attache pour faire baisser la fréquence ou au niveau de l'extrémité pour faire monter la fréquence.

Quant à la solution physique, elle est toujours obtenue par addition des solutions harmoniques, mais elle ne produit plus en général un signal périodique. C'est un son que l'on peut juger moins « mélodieux». Toutefois, il faut prendre en compte le résonateur ou l'oreille qui peuvent éliminer les partiels élevés, ce qui est le cas par exemple des boîtes à musique.

2.3 Les membranes

Outre les tiges frappées, la section des percussions comporte aussi des instruments constitués d'une membrane élastique fixée sur un pourtour circulaire : timbales, tambours, tam-tams... Considérons une telle membrane de rayon R, disposée horizontalement, et dont la hauteur est une fonction $u(x, y, t)$ telle que $u(x, y, t) = 0$ si $r = \sqrt{x^2 + y^2} = R$ (le pourtour).

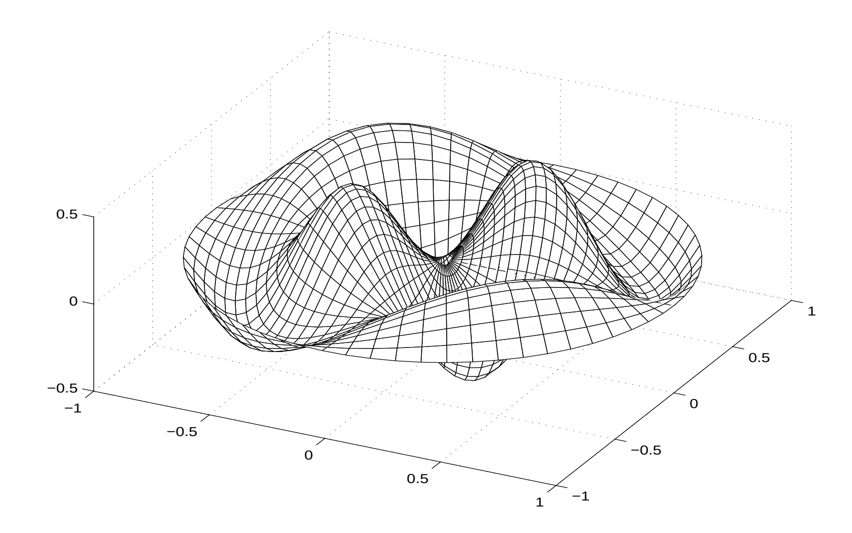


FIG. 2.7 – huitième mode propre d'une timbale.

On suppose que la tension T est uniforme dans toute la membrane, et la masse par unité de surface est notée μ. En reprenant la même démarche que celle suivie pour la corde (légèrement compliquée par le passage en dimension 2), on peut montrer que u est solution de l'équation des ondes (avec $\Delta = \partial_{x^2} + \partial_{y^2}$) :

$$\mu \partial_{t^2} u(x, y, t) = T \Delta u(x, y, t). \tag{2.14}$$

Les solutions harmoniques de fréquence f, de la forme $u(x, y, t) = \varphi(x, y) \exp(2i\pi f t)$, donnent reportées dans (2.14), toujours avec $k = 2\pi f/c$ et $c = \sqrt{T/\mu}$:

$$\Delta\varphi + k^2\varphi = 0.$$

Le bord de la membrane étant circulaire, il est commode pour l'analyse de passer en coordonnées polaires (r, θ). Cherchons les solutions qui peuvent s'écrire sous la forme séparée $\varphi(x, y) = \gamma(r)\sigma(\theta)$, en utilisant la formule du laplacien en coordonnées polaires $\Delta = \partial_{r^2} + \partial_r/r + \partial_{\theta^2}/r^2$. Après division par $\gamma(r)\sigma(\theta)/r^2$, le calcul donne

$$r^2 \frac{\gamma''(r)}{\gamma(r)} + r\frac{\gamma'(r)}{\gamma(r)} + k^2 r^2 = -\frac{\sigma''(\theta)}{\sigma(\theta)}.$$

Cette expression est nécessairement une constante que nous notons m^2. On a alors d'une part $\sigma''(\theta) = -m^2\sigma(\theta)$, d'où

$$\sigma(\theta) = \alpha \exp(im\theta) + \beta \exp(-im\theta).$$

De plus, la fonction σ étant 2π-périodique, m est nécessairement un entier. On a d'autre part

$$r^2\gamma''(r) + r\gamma'(r) + (k^2r^2 - m^2)\gamma(r) = 0,$$

qui est *l'équation de Bessel*, dont les solutions bornées en zéro sont les fonctions de Bessel d'ordre m de première espèce, notées J_m :

$$\gamma(r) = \alpha J_m(|k|r).$$

Il reste à prendre en compte la condition aux limites (sur le pourtour) $\gamma(R) = 0$. Cela impose $J_m(|k|R) = 0$, donc $|k|R$ est un zéro de J_m. Les zéros de J_m sont notés z_{mn} :

$$J_m(z_{mn}) = 0.$$

Le tableau 2.1 donne les premières valeurs de z_{mn}.

$m\backslash n$	0	1	2	3	4	5
0	-	2.40	5.52	8.65	11.79	14.93
1	0	3.83	7.02	10.17	13.32	16.47
2	0	5.14	8.42	11.62	14.80	17.96
3	0	6.38	9.76	13.02	16.22	19.41
4	0	7.59	11.06	14.37	17.62	20.83
5	0	8.77	12.34	15.70	18.98	22.22

TAB. 2.1 – zéros de J_m.

Les solutions harmoniques sont donc de la forme

$$u(x, y, t) = J_m(z_{mn}r/R)\,[\alpha \exp(im\theta) + \beta \exp(-im\theta)] \exp(\pm 2i\pi f_{mn}t)$$

avec les fréquences propres

$$f_{mn} = \frac{z_{mn}c}{2\pi R}, \tag{2.15}$$

et les solutions de l'équation des ondes s'obtiennent ici encore par superposition de ces solutions harmoniques.

Contrairement aux cas des cordes ou des lames, ces ondes ne sont pas nécessairement stationnaires. Pour $m \geq 1$, il peut y avoir des ondes tournantes de la forme (ici dans le sens de l'horloge)

$$u(x, y, t) = \alpha J_m(z_{mn}r/R) \exp(2i\pi f_{mn}t - im\theta),$$

la vitesse angulaire de rotation étant $2\pi f_{mn}/m$: la valeur de u est constante sur tout point mobile décrivant un cercle de centre $(0, 0)$ et d'équation $\theta - 2\pi f_{mn}t/m =$ constante.

Les modes propres stationnaires sont quant à eux de la forme

$$u(x, y, t) = \alpha J_m(z_{mn}r/R) \sin(m(\theta - \theta_0)) \exp(\pm 2i\pi f_{mn}t).$$

Sur la figure 2.8 sont représentées les lignes de niveau des huit premiers modes. Les valeurs z_{mn} ont été calculées numériquement à partir des *valeurs propres* de la matrice d'éléments finis [9] associée au problème, et sont proches de celles données dans le tableau 2.1.

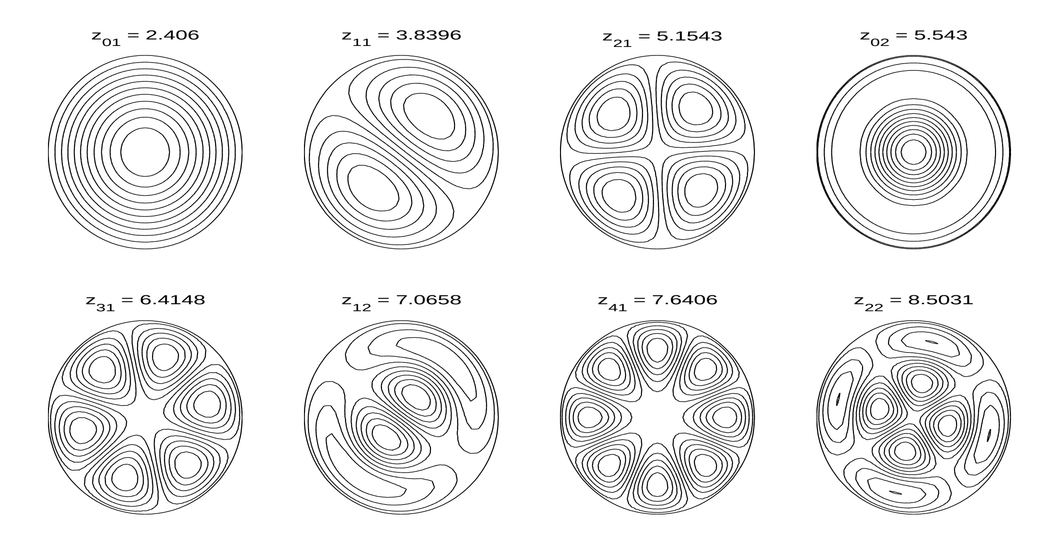


FIG. 2.8 – lignes de niveau des huit premiers modes propres calculés numériquement.

2.4 Les tubes

Quittons maintenant les percussions pour rejoindre la section des vents, et examinons le résonateur de ces instruments : le tube. Contrairement aux cas abordés jusqu'à présent, il s'agit d'instruments de musique *à son entretenu*, soit directement par le souffle de l'instrumentiste (bois, cuivres), soit par une soufflerie mécanique (orgue). Nous étudions ici le cas d'un tube *cylindrique* de longueur L (cf. fig. 2.9), en se plaçant dans l'hypothèse où il est parcouru par une onde plane[4], de direction l'axe du tube Ox. La pression acoustique p_a dans le tube ne dépend alors que de x et de t, et on la note simplement $p(x,t)$. La vitesse (moyenne) des particules d'air dans le tube est notée $v(x,t)$. Dans le modèle simplifié que nous décrivons, l'excitation acoustique produite par l'embouchure est donnée, et l'on étudie la réaction du tube à cette excitation. On peut alors distinguer deux types d'excitation ou de « commande» du phénomène acoustique.

- **Commande en pression :** dans ce cas, la source des vibrations de l'air consiste en une pression $p_0(t)$ imposée à l'entrée (ici à gauche) du tube : $p(0,t) = p_0(t)$ pour tout t. Ce modèle correspond approximativement aux vents dépourvus d'anche, comme par exemple la flûte ou les « tuyaux à bouche» de l'orgue (qui fonctionnent sur le principe de la flûte).
- **Commande en vitesse :** dans ce cas, c'est la vitesse de l'air $v_0(t)$ qui est imposée à l'entrée du tube : $v(0,t) = v_0(t)$ pour tout t. Ce modèle correspond approximativement aux instruments à anche, comme par exemple la clarinette ou les « tuyaux à anche» de l'orgue.

[4]Cette hypothèse est valable si l'on suppose que la paroi du tube est parfaitement rigide et étanche. Dans ce cas, la vitesse perpendiculaire à la paroi est nulle sur la paroi, et l'on déduit de l'équation d'Euler que la dérivée normale de la pression est nulle sur la paroi : $\nabla p.\mathbf{n} = 0$, $\mathbf{n}$ étant un vecteur unitaire perpendiculaire (« normal») à la paroi. Ceci rend possible le fait qu'à t fixé, la pression soit constante en tout point d'une section droite du tube, autrement dit soit une onde plane (il en existe cependant d'autres types).

Comme nous allons le voir, le type d'excitation ou commande a, dans le cas d'un tube cylindrique, une influence majeure sur le timbre : présence ou absence des harmoniques de rang impair, mais également hauteur du fondamental. À noter que cette différence s'estompe, voire disparaît, dans le cas d'un tube conique (hautbois, basson, saxophone) comme le montrent les sujets d'étude 2.6.10 et 2.6.11. Le terme de *commande* (en pression ou en vitesse) fait ici référence à la condition qui est imposée à l'entrée du tube, et non à ce que peut éprouver le musicien qui n'agit pas directement sur le tube, mais sur ce qui le met en vibration : anche, filet d'air, lèvres (cf. aussi remarque 2.2). Par ailleurs, il existe encore d'autres types de commandes *mixtes* (cf. § 2.6.9) où ce qui est imposé concerne $\alpha\, p(0,t) + \beta\, v(0,t)$, et dont les précédentes ne sont que des cas particuliers.

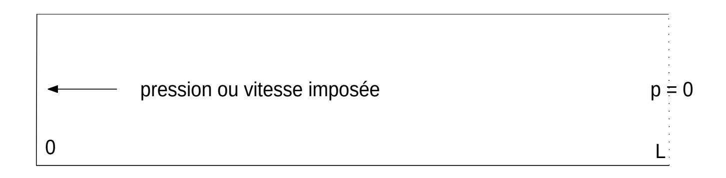


FIG. 2.9 – pression dans un tube cylindrique (coupe) avec les conditions aux limites.

En tout état de cause, il s'agit là de deux modèles simplifiés, et la réalité est plus complexe. En particulier, les tubes ne sont pas forcément cylindriques ou coniques : ils peuvent s'évaser à leur extrémité comme le pavillon des cuivres ; la source et le tube peuvent interagir de manière non linéaire, rendant l'étude plus difficile, et faisant encore l'objet de recherche intensives. De plus, nous supposons que le tube n'a pas d'autres ouvertures que les extrémités, ce qui est loin de la réalité : de nombreux instruments à vent sont percés de trous latéraux pour pouvoir changer de note ! Et considérer que la présence d'un trou donne à peu près le même résultat que si le tube était coupé à son niveau n'est qu'une approximation assez grossière.

2.4.1 Commande en pression

La source des vibrations est donc ici la pression à l'entrée du tube $p_0(t)$. Par ailleurs, une approximation raisonnable est obtenue en supposant que la pression acoustique est nulle à la sortie du tube (tuyau ouvert sur l'air environnant). Ce n'est pas tout à fait exact, et les facteurs d'instruments à vent savent en tenir compte en modifiant en conséquence la longueur du tuyau (ou la position des trous) par rapport à la longueur donnée par le modèle simplifié. Compte tenu de l'étude de l'équation des ondes menée au chapitre 1, nous obtenons le système linéaire suivant :

$$\begin{cases} \partial_{t^2}p(x,t) - c^2\partial_{x^2}p(x,t) = 0, & \text{dans le tube,} \\ p(0,t) = p_0(t), & \text{à l'entrée,} \\ p(L,t) = 0, & \text{à la sortie.} \end{cases} \tag{2.16}$$

La source $p_0(t)$ est supposée être périodique de période $T = 1/f$. Elle peut donc être décomposée en série de Fourier

$$p_0(t) = \sum_{n=-\infty}^{+\infty} c_n \exp(2i\pi f_n t), \quad f_n = nf.$$

Le système (2.16) étant linéaire, la réponse du tube sera la somme des réponses à chacune des excitations $c_n \exp(2i\pi f_n t)$, et l'on peut donc restreindre l'étude au cas où l'excitation est harmonique, c'est-à-dire

$$p_0(t) = \exp(2i\pi f t), \tag{2.17}$$

ce que nous supposons désormais. La pression est alors aussi harmonique, c'est-à-dire de la forme

$$p(x,t) = \varphi(x)\exp(2i\pi f t).$$

En reportant cette expression dans (2.16), on retrouve que $\varphi(x)$ est solution de l'équation de Helmholtz (où $k = 2\pi f/c$), avec des conditions aux limites non homogènes :

$$\begin{cases} \varphi''(x) + k^2\varphi(x) = 0, & \text{dans le tube,} \\ \varphi(0) = 1, & \text{à l'entrée,} \\ \varphi(L) = 0, & \text{à la sortie.} \end{cases} \tag{2.18}$$

La solution générale de la première équation est $\varphi(x) = \alpha\exp(ikx) + \beta\exp(-ikx)$, et les conditions aux limites imposent

$$\begin{cases} \alpha + \beta = 1, \\ \alpha\exp(ikL) + \beta\exp(-ikL) = 0. \end{cases}$$

Ce système linéaire a une solution unique si et seulement si le déterminant est non nul, autrement dit

$$\sin(kL) \neq 0. \tag{2.19}$$

Si tel est le cas, le calcul donne $\alpha = i\exp(-ikL)/(2\sin(kL))$, $\beta = -i\exp(ikL)/(2\sin(kL))$, et l'on a

$$\varphi(x) = \frac{i\exp(-ik(L-x)) - i\exp(ik(L-x))}{2\sin(kL)} = \frac{\sin(k(L-x))}{\sin(kL)}.$$

La pression dans le tube soumis à l'excitation harmonique (2.17) est donc

$$p(x,t) = \frac{\sin(k(L-x))}{\sin(kL)}\exp(2i\pi f t). \tag{2.20}$$

La première remarque que l'on peut faire est que l'on obtient une onde stationnaire, comme pour les cordes et les lames. Maintenant, que se passe-t-il pour les valeurs « interdites» où $\sin(kL) = 0$?

Pour répondre à cette question, examinons en particulier ce qui se passe à la sortie du tube en $x = L$, où est produit le son qui va rayonner dans l'air libre. La pression y est constamment nulle, c'est donc plutôt la vitesse que nous allons considérer. D'après l'équation d'état (1.5), nous savons que

$$c^2\rho_0\partial_x v(x,t) = -\partial_t p(x,t) = -2i\pi f\frac{\sin(k(L-x))}{\sin(kL)}\exp(2i\pi f t).$$

En intégrant cette équation par rapport à x, on obtient

$$v(x,t) = \frac{\cos(k(L-x))}{ic\rho_0\sin(kL)}\exp(2i\pi f t) + g(t). \tag{2.21}$$

Pour déterminer la valeur de la constante (par rapport à x) d'intégration $g(t)$, on utilise l'équation d'Euler (1.3). On en déduit que $g'(t) = 0$, et si l'on suppose que la vitesse est de moyenne nulle, on a $g = 0$. Ainsi, en passant dans le domaine physique (*i.e.,* en prenant les parties réelles), à une *pression d'entrée* de fréquence f

$$p_0(t) = \cos(2\pi f t),$$

correspond *une vitesse de l'air en sortie* de tube

$$v(L,t) = \frac{1}{c\rho_0 \sin(kL)} \sin(2\pi f t),$$

qui sera d'autant plus grande que $\sin(kL)$ est proche de 0, et théoriquement infinie si $\sin(kL) = 0$, c'est-à-dire si $k = n\pi/L$.

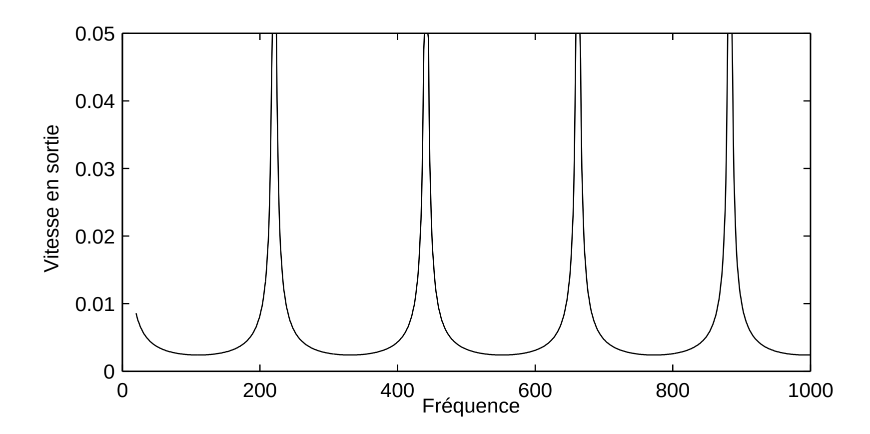


FIG. 2.10 – vitesse absolue en sortie du tube en fonction de la fréquence. Le tube entre en résonance aux fréquences où apparaissent les pics.

En réalité, il y a un léger amortissement dû à une dissipation de l'énergie sous forme de chaleur, et cette vitesse en sortie ne sera pas infinie, mais seulement très grande. Les fréquences propres associées à ces valeurs de k,

$$f_n = \frac{nc}{2L} = n f_1, \quad n \geq 1$$

produiront donc un son puissant et seront favorisées au détriment des autres (cf. fig. 2.10) : ce sont celles-ci qui se produiront lorsque le musicien soufflera dans l'instrument. Leur progression étant proportionnelle à la suite des entiers

$$1, 2, 3, \ldots, n, \ldots$$

nous sommes en présence d'harmoniques.

Pour l'auditeur, le son résultant $s(t)$ sera alors de la forme

$$s(t) = \sum_{n \geq 1} \alpha_n \sin(2n\pi f_1 t + \theta_n),$$

son périodique de hauteur (le fondamental) donnée par la loi de Bernoulli :

$$f_1 = \frac{c}{2L}.$$

On remarque que c'est la même formule que celle qui nous a donné le fondamental de la corde (mais pas le même c). Par exemple, pour qu'un tuyau d'orgue donne le Do grave à 32.7Hz, il faut un tuyau de longueur $L = 340/(2\times\ 32.7) = 5.2$m !

Pour les vents, l'accord se fait (mise à part la facture) par ajustement de la longueur du tuyau au niveau de l'embouchure. Pour les orgues, il se fait en déplaçant une bague coulissante à l'extrémité du tuyau (fig. 2.11), mais également au niveau de l'anche en jouant sur sa longueur (cas des lames) ou au niveau de la sortie en jouant sur sa forme plus ou moins ouverte. Sachant qu'un orgue peut comporter plusieurs milliers du tuyaux (l'orgue de l'opéra de Sydney en compte 10500), cela représente un travail conséquent !

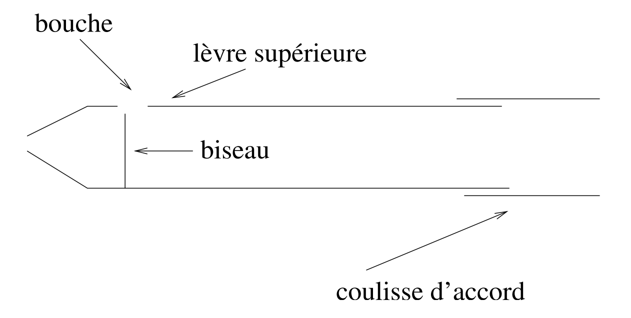


FIG. 2.11 – accord par ajustage de la longueur.

Remarque 2.2 *Si l'on examine le cas limite où* $\sin(kL) = 0$, *c'est-à-dire* $kL = n\pi$, *en* multipliant *par* $\sin(kL)$ *les équations (2.20) et (2.21), on obtient les fonctions*

$$p(x,t) = \pm \sin(n\pi x/L)\exp(2i\pi f t),$$
$$v(x,t) = \mp\frac{\cos(n\pi x/L)}{ic\rho_0}\exp(2i\pi f t),$$

qui sont encore solution de l'équation des ondes, mais plus des conditions aux limites (2.16). En particulier, la pression est nulle à l'entrée du tube alors que la vitesse y atteint son amplitude maximale. Cela fait dire à certains auteurs [20] qu'il s'agit là d'une commande en vitesse et non en pression. Ces modes correspondent cependant aux fréquences pour lesquelles le problème de la commande en pression s'est trouvé être singulier.

2.4.2 Commande en vitesse

Dans les instruments à anche, c'est plutôt la vitesse de l'air à l'entrée du tube qui commande la résonance du tube. L'anche agit comme une soupape, alternativement ouverte ou partiellement fermée au passage de l'air, selon qu'elle est plus ou moins décollée du bec. À

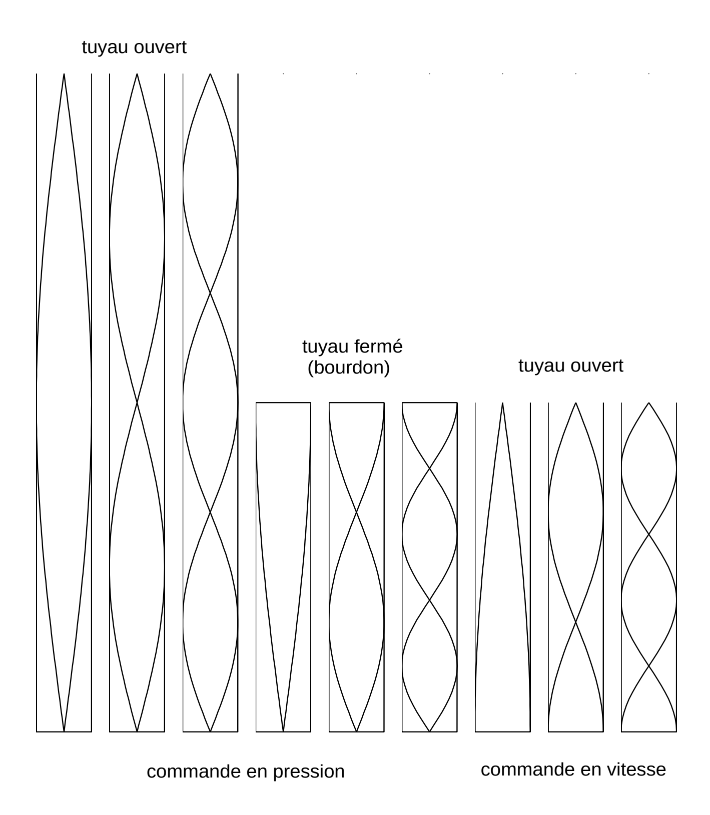


FIG. 2.12 – noeuds et ventres de la pression pour les trois premiers modes, en fonction de la commande et de la nature de la sortie, ouverte ou fermée. Les tuyaux deux fois plus courts produisent le même fondamental, mais n'ont pas d'harmoniques pairs.

noter qu'un haut débit fourni par l'instrumentiste aura tendance à fermer l'admission d'air, au contraire de ce qui se passe au niveau des lèvres du trompettiste.

La vitesse obéit comme la pression à l'équation de Helmholtz. Dans le cas harmonique, pour une vitesse d'entrée $v_0(t) = \exp(2i\pi ft)$, elle est de la forme $v(x,t) = \varphi(x)\exp(2i\pi ft)$ avec

$$\begin{cases} \varphi''(x) + k^2\varphi(x) = 0, & \text{dans le tube,} \\ \varphi(0) = 1, & \text{à l'entrée,} \\ \varphi'(L) = 0, & \text{à la sortie.} \end{cases} \tag{2.22}$$

La condition en sortie de tube $\varphi'(L) = 0$, appelée condition de Neumann, provient de l'équation d'état (1.5) et de la condition $p(L,t) = 0$, qui donnent

$$c^2\rho_0\partial_x v(L,t) = -\partial_t p(L,t) = 0,$$
$$c^2\rho_0\varphi'(L)\exp(2i\pi ft) = 0.$$

La solution générale de la première équation de (2.22) est $\varphi(x) = \alpha\exp(ikx) + \beta\exp(-ikx)$, et les conditions aux limites imposent maintenant

$$\begin{cases} \alpha + \beta = 1, \\ \alpha ik\exp(ikL) - \beta ik\exp(-ikL) = 0. \end{cases}$$

Ce système linéaire a une solution unique si et seulement si (comparer avec (2.19))

$$\cos(kL) \neq 0.$$

Si tel est le cas, le calcul donne $\alpha = \exp(-ikL)/(2\cos(kL))$ et l'on obtient

$$\varphi(x) = \frac{\cos(k(L-x))}{\cos(kL)}.$$

La vitesse dans le tube est donc

$$v(x,t) = \frac{\cos(k(L-x))}{\cos(kL)} \exp(2i\pi ft),$$

et à une *vitesse d'entrée* $v(0,t) = \cos(2\pi ft)$ correspond *une vitesse en sortie*

$$v(L,t) = \frac{1}{\cos(kL)} \cos(2\pi ft).$$

Les valeurs critiques ont changé ! Ce ne sont plus les fréquences telles que $\sin(kL) = 0$ qui vont être amplifiées, mais celles pour lesquelles $\cos(2\pi fL/c) = \cos(kL) = 0$, autrement dit les fréquences propres

$$f_n = \frac{(n-1/2)c}{2L}, \quad n \geq 1.$$

Nous pouvons faire alors deux constatations intéressantes.

- La première fréquence propre est

$$f_1 = \frac{c}{4L},$$

  la moitié de celle que nous avions observée pour la commande en pression. Sous ce régime de fonctionnement, l'instrument joue une octave en-dessous ! Le même phénomène se produit d'ailleurs dans le cas d'un tube commandé en pression, *fermé à l'autre extrémité* (cf. § 2.6.8). Dans les orgues, de tels tuyaux s'appellent des *bourdons* (cf. fig. 2.12).
- La suite des fréquences propres est donnée par $f_n = (2n-1)f_1$: leur progression est maintenant proportionnelle aux entiers *impairs*

$$1, 3, 5, \ldots, 2n-1, \ldots,$$

  les harmoniques pairs ont disparu ! Cette absence est justement l'un des éléments qui permettent à l'auditeur de reconnaître des instruments à anche comme la clarinette, et leur donne cette sonorité que certains qualifient de « nasillarde». Elle explique également le fait que la clarinette « quintoye» : lorsque l'on souffle plus fort, on passe directement du registre fondamental (registre de chalumeau) au registre situé une octave plus une quinte au-dessus (registre de clairon), alors que sur une flûte, on passe seulement à l'octave supérieure.

En superposant les différents modes harmoniques, on obtient un son résultant

$$s(t) = \sum_{n \geq 1} \alpha_n \cos[(2n-1)\pi f_1 t + \theta_n],$$

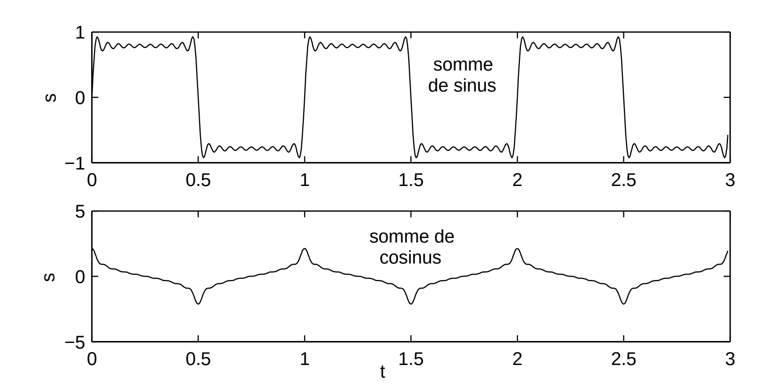


FIG. 2.13 – malgré les apparences, ces deux signaux sont composés exactement des mêmes fréquences ! Ce sont les phases des harmoniques qui les distinguent.

son périodique de hauteur f_1. Rappelons qu'une telle somme peut s'écrire indifféremment avec des sinus ou des cosinus dès lors que l'on fait apparaître les phases θ_n. Les phases peuvent d'ailleurs avoir une incidence notoire sur l'allure du son. La figure 2.13 reproduit respectivement les deux sons

$$s_1(t) = \sum_{n=1}^{10} \frac{1}{2n-1} \sin((2n-1)\pi t),$$
$$s_2(t) = \sum_{n=1}^{10} \frac{1}{2n-1} \cos((2n-1)\pi t).$$

Graphiquement, la différence semble importante, et pourtant, ce sont exactement les mêmes fréquences qui sont présentes. Cependant, l'audition de ces deux signaux ne révèle que peu de différences : l'oreille semble assez peu sensible à la phase. Dans le cas de la clarinette, on peut observer l'une ou l'autre de ces deux formes (parmi d'autres intermédiaires), selon la hauteur et l'intensité de la note jouée [4].

2.5 Timbre des instruments

Définir ce qu'est le *timbre* d'un instrument n'est pas une chose simple, et la littérature sur le sujet est aussi abondante que diversifiée. Nous nous contentons pour le moment de décrire et d'illustrer par quelques exemples deux caractéristiques permettant de distinguer entre eux (de manière certes incomplète) les divers instruments :

- la nature du spectre du son, qui dépend d'une part de l'excitateur (corde, anche, lame, membrane), et d'autre part du résonateur (table d'harmonie, tuyau) qui va amplifier et « colorer» le son produit,
- l'enveloppe du son, qui définit la manière dont naît, vit et meurt un son musical donné.

Il y a bien d'autres éléments qui rentrent en ligne de compte comme le vibrato, le souffle dans les vents ou l'impact initial dans les instruments à percussion, la réverbération des

autres cordes dans un piano, ou encore le déphasage (effet Doppler, effet Leslie) dont usent par exemple les musiciens de jazz en bougeant ou en faisant tourner leur instrument. Les gambistes aussi ont cette étrange coutume, qui leur donne paraît-il le mal de mer ! Nous aborderons quelques uns de ces aspects au chapitre 6.

2.5.1 Nature du spectre

Tous les instruments de musique produisent des sons qui ont en gros un spectre discret. Par conséquent, les décrire revient à dire quelles sont les fréquences f_n présentes, et quelles sont la phase θ_n et l'amplitude α_n relatives à ces fréquences, dans la représentation du son

$$s(t) = \sum_{n\geq 1} \alpha_n \cos(2\pi f_n t + \theta_n),$$

valable au moins sur un intervalle de temps assez court. Ces trois ensembles de données donnent déjà lieu à une grande diversité de timbres.

Harmoniques ou partiels, inharmonicité du piano

Une première caractéristique que l'oreille d'un musicien sait parfaitement distinguer est si le son est périodique ou non : il ne confondra pas le son d'un piano avec celui d'une cloche, c'est-à-dire des harmoniques (rappelons que dans ce cas tous les f_n sont des multiples entiers de f_1) avec des partiels (les autres cas), alors que ces instruments sont tous deux à percussion et ont une enveloppe assez ressemblante, caractérisée par un impact suivi d'une décroissance marquée.

Cependant, certains pianos de faibles dimensions (petits pianos droits et crapauds), aux cordes plus courtes compensées par un diamètre plus fort[5], produisent un son un peu acide qui justement fait penser à un son de cloche. Cela provient précisément du fait que les fréquences produites par une telle corde s'écartent de la progression des entiers chère aux pythagoriciens. C'est ce que l'on appelle *l'inharmonicité*. Une façon élémentaire d'expliquer sa cause est la suivante : le deuxième harmonique est produit par la division de la corde en deux longueurs égales, deux ventres séparés par un nœud (fig. 2.2). Ceci vaut pour une corde parfaite. Mais pour une corde réelle, qui présente une certaine raideur, le nœud ne se réduit pas à un point, ce qui fait que les longueurs des deux ventres s'en retrouvent légèrement réduites, et a pour effet d'augmenter un peu la fréquence du second harmonique, et ce d'autant plus que la corde sera plus courte. Le même phénomène, accentué, se produit pour les harmoniques supérieurs.

Pour mesurer l'inharmonicité, on utilise comme unité logarithmique le *cent* ou centième de demi-ton. Un demi-ton tempéré correspond à un rapport de fréquences égal à $\sqrt[12]{2} \simeq 1.05946$, et le cent correspond par conséquent à un rapport de fréquences $\sqrt[1200]{2} \simeq 1.0005778$. Dire que deux fréquences $f_1 < f_2$ sont à 1 cent d'intervalle signifie que

$$\log_2 f_2 = \log_2(f_1 \times 2^{1/1200}) = \log_2 f_1 + \frac{1}{1200} = \log_2 f_1 + 1\,\text{cent},$$

[5]La loi de Taylor $f = \sqrt{T/\mu}/(2L)$ nous dit que l'on ne change pas la fréquence si l'on maintient constante la valeur $L\sqrt{\mu}$, où μ est la masse linéique, proportionnelle au carré du diamètre. Par exemple, si l'on divise par 2 la longueur et multiplie par 2 le diamètre sans modifier la tension, on obtient la même note. Le problème est que l'on a augmenté la raideur de la corde.

où $\log_2$ est le logarithme en base 2 ($\log_2 2^n = n$), commode à utiliser dans la mesure où l'intervalle d'octave correspond justement à un rapport 2. Les facteurs de piano ont alors constaté que pour une note de fondamental théorique f, l'écart en cents d'un partiel par rapport à l'harmonique correspondant est à peu près proportionnel au carré du rang de l'harmonique :

$$\log_2 f_n \simeq \log_2(nf) + \frac{\zeta(f)n^2}{1200}.$$

La valeur de $\zeta(f)$ dépend de la note et de l'instrument. Dans [7], on trouve la formule

$$\zeta(f) \simeq \frac{d^2}{L^4 f^2} 3.3\,10^9$$

où d est le diamètre de la corde, L sa longueur et f sa fréquence, ce que confirme l'étude proposée au § 2.6.6. Cela donne par exemple $\zeta(f) \simeq 0.3$ dans le médium d'un piano de 107cm de haut. Le 10ième partiel se retrouve alors 30 cents au-dessus de l'harmonique correspondant, soit un tiers de ton ! Sur un piano à queue avec une corde de diamètre divisé par $\sqrt{2}$ mais deux fois plus longue (mais aussi avec une tension deux fois plus forte pour conserver la même fréquence), cette valeur sera divisée par $2 \times 2^4 = 32$, et l'écart ne sera plus que d'environ 1 cent. Tout pianiste entend parfaitement la différence, même s'il ignore son origine.

Richesse en harmoniques aigus

Une seconde caractéristique du timbre est la répartition des harmoniques (ou des partiels) : un son riche en harmoniques aigus sera qualifié de « brillant», voire « métallique», alors qu'un son pauvre en harmoniques aigus sera qualifié de « rond», « doux», voire « terne». Dans la première catégorie, on trouve par exemple le violon, le clavecin, un piano aux marteaux très durs. Dans la seconde, on trouve les flûtes, certains jeux d'orgues, un piano avec des marteaux au feutre mou. Les figures 2.14 et 2.15 montrent respectivement le son d'une flûte traversière et d'un clavecin [16]. La différence de richesse en harmoniques aigus est frappante. La classification par instruments est cependant discutable car un même instrument, selon la manière dont on en joue, et en particulier selon l'intensité, produira plus ou moins d'harmoniques aigus : sur à peu près tous les instruments, on constate que l'intensité *relative* des harmoniques aigus (le rapport $|c_n/c_1|$) augmente avec la force avec laquelle est jouée la note, ce qui traduit typiquement un comportement *non linéaire* de l'instrument.

Une manière d'analyser la répartition des harmoniques d'un son consiste à observer à quelle vitesse les coefficients de Fourier (les amplitudes des harmoniques) décroissent. Un résultat mathématique énonce que si une fonction périodique admet des dérivées de carré intégrable jusqu'à l'ordre m, alors ses coefficients de Fourier c_n vérifient

$$\sum_{n=-\infty}^{\infty} |n^m c_n|^2 < \infty,$$

et en particulier décroissent plus rapidement que $1/n^m$: le terme général de la série tendant vers 0, on a

$$|c_n| = \frac{\varepsilon(n)}{n^m} \text{ où } \lim_{n\to\infty} \varepsilon(n) = 0.$$

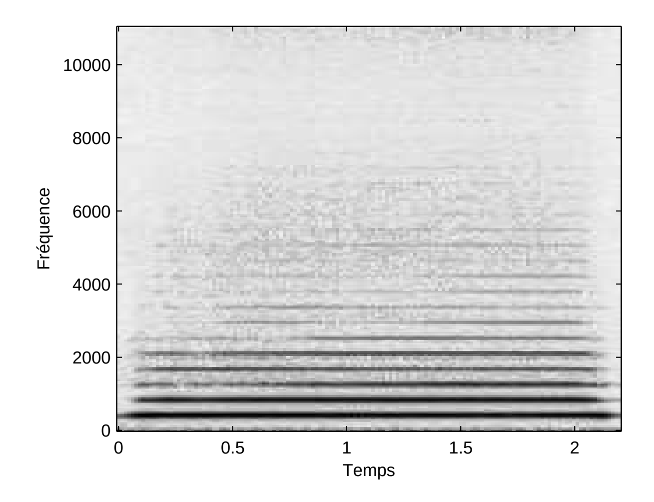


FIG. 2.14 – harmoniques d'une flûte traversière (Sol♯-3) à quatre petites clés.

Or, si l'on reprend l'exemple de la corde libre, frappée ou pincée (cf. (2.4), (2.5), (2.6), ainsi que § 2.6.2), on constate que les coefficients de Fourier sont déterminés par les conditions initiales. À des conditions initiales « peu dérivables» - pincement par le bec du sautereau du clavecin, avec l'ongle ou un médiator pour la guitare - correspondront des coefficients de Fourier à décroissance relativement lente, d'où un son riche en harmoniques aigus, et vice-versa.

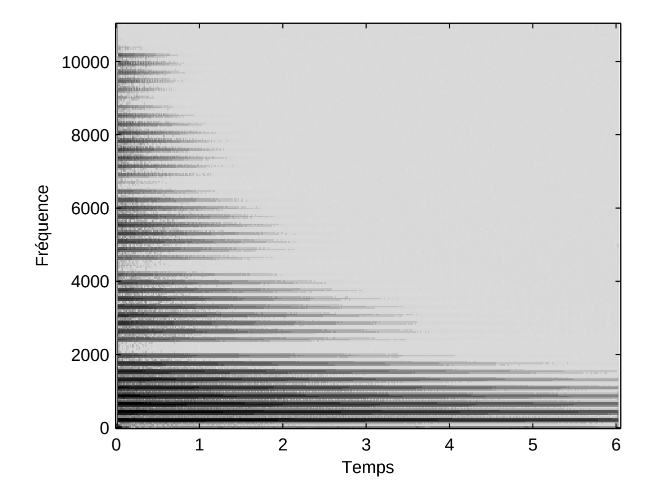


FIG. 2.15 – note de clavecin (La-2) pincée au dixième de sa longueur, ce qui a pour effet de supprimer les harmoniques de rang multiple de 10.

De même, dans le cas du violon, l'archet impose à la corde un mouvement en dents de scie, appelé signal triangulaire, représenté schématiquement sur la figure 2.16 (cf. aussi § 2.6.3). À chaque période, la corde est d'abord entraînée par l'archet jusqu'à ce que la tension de la

corde l'emporte sur les forces de frottement, provoquant un brusque retour de la corde. Un tel signal correspond au cas $m = 1$, a des coefficients en $1/n^2$, et le cas limite où la pente du « retour» serait infinie (signal discontinu) donnerait des coefficients en $1/n$. C'est une décroissance relativement peu rapide, d'où ici encore un son riche en harmoniques aigus.

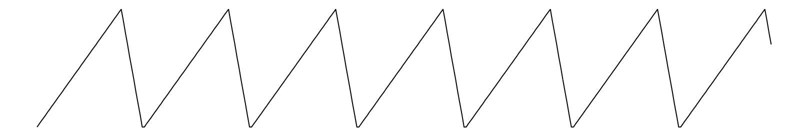


Fig. 2.16 – quelle scie...

Répartitions diverses des harmoniques

Nous avons déjà vu que dans le cas périodique, une autre caractéristique qui ne trompe pas est la quasi absence d'harmoniques pairs. C'est le cas en particulier de la clarinette et du bourdon de l'orgue. De manière analogue, un grand nombre de timbres différents peut être obtenu en favorisant une certaine catégorie d'harmoniques par rapport aux autres. Cette technique est employée de manière intensive par les facteurs d'orgue.

Dans la *synthèse des principaux*[6][19], ils additionnent *plusieurs tuyaux* pour jouer une même note, Do-1 par exemple, *correspondant à la progression des harmoniques* : Do-1 (principal), Do-2 (prestant), Sol-2 (quinte), Do-3 (doublette), Mi-3 (tierce), Sol-3 (quinte), Do-4 (sifflet)... On remarque l'absence du septième harmonique, jugé peu esthétique. En continuant d'enrichir avec des octaves, tierces et quintes supérieures, on obtient des jeux aux nom évocateurs : fournitures, cymbales, mixtures, plein-jeux. À noter que lorsqu'on voulut construire tout un clavier sur ce principe, se posa le problème qu'à partir d'une certaine hauteur, il n'était plus possible de raccourcir les tuyaux ! Les facteurs d'orgue utilisèrent alors la *reprise*, consistant à décaler vers le bas d'une octave les notes ajoutées, allant jusqu'à superposer des graves aux notes aiguës ! Nous verrons d'ailleurs au chapitre 4 que cette technique constitue l'ingrédient du « son ascendant perpétuel» de Shepard (son qui semble monter indéfiniment, tout en repassant périodiquement par exactement les mêmes notes). Dans les *synthèses flûtées* où sont mis en œuvre des tuyaux de taille plus large, aux sons plus doux, le principe est le même, sauf que la synthèse est limitée aux six premiers harmoniques, avec présence obligatoire du cinquième (jeux de tierce et de cornet). Enfin, dans les *synthèses d'anches* n'interviennent que des octaves.

Rôle du résonateur

Tous les instruments de musique utilisent un résonateur pour rayonner efficacement les vibrations produites par l'instrumentiste, même la rudimentaire guimbarde qui utilise la résonance de la bouche. Pour les instruments à corde, le résonateur est une table d'harmonie en épicéa, qui reçoit l'énergie mécanique de la corde par l'intermédiaire du chevalet

[6] Les principaux sont des jeux à bouche de « taille» (diamètre) intermédiaire entre les jeux à taille large (flûtes et bourdons) de son assez doux, et les jeux à taille étroite (gambes et saliciaux) conçus dans l'intention d'imiter les cordes.

(fig. 2.17). Pour les vents, le résonateur est le tube lui-même. Pour les percussions, c'est en général une caisse.

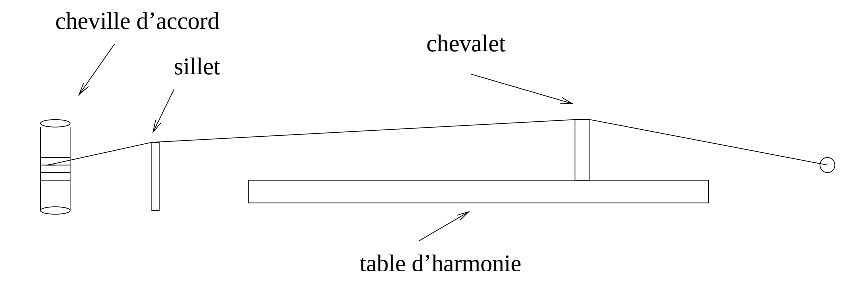


FIG. 2.17 – l'inséparable corde-résonateur.

Dans tous les cas, on constate en première approximation que la réponse du résonateur à l'excitation est linéaire : par exemple la réponse du violon à deux cordes jouées simultanément est égale à la somme de celles des cordes jouées séparément. On constate également qu'elle est invariante par rapport au temps (cf. (1.19)) : de manière imagée, écouter demain ce que vous jouez aujourd'hui revient au même qu'attendre un jour pour écouter ce que vous jouerez demain. Ceci paraît anodin (voire idiot), et pourtant, ces deux seules hypothèses permettent de dire que *le résonateur est un filtre* (cf. § 1.5), amplifiant certaines fréquences, en atténuant d'autres, mais sans jamais faire changer la fréquence d'un harmonique donné. Par exemple, la courbe de la figure 2.10 n'est rien d'autre que la fonction de transfert d'un instrument à vent. On pourrait dire que l'art du facteur d'instrument consiste en grande partie à « ajuster» la fonction de transfert au goût des musiciens. Parfois, celle-ci peut être modifiée par le musicien lui-même, comme c'est le cas pour le pavillon de la trompette qu'il bouche avec la main ou une sourdine.

Ainsi, comme nous l'avions constaté sur l'exemple de la figure 2.10, ce sont *les fréquences propres du résonateur qui seront amplifiées.* Parfois, cela a des effets indésirables, comme ce qui se passe sur des instruments qui ont un mode propre trop prononcé. Par exemple, la première résonance du violon, qui est une résonance de la cavité, se situe à peu près au niveau du Do♯, et est suivie juste au-dessus par un « trou» dans la réponse [4], au niveau du Ré. Ainsi, en notant f_1 la fréquence du Do♯, f_2 celle du Ré et $H(f)$ la fonction de transfert du résonateur, on a dans ce cas $|H(f_2)| \ll |H(f_1)|$. En supposant que le signal $e(t)$ fourni par le chevalet soit sinusoïdal de fréquence f, le son produit par le violon sera $H(f)e(t)$ (cf. § 1.5). Si l'on joue les deux notes consécutivement, le Do♯ sonnera alors bien plus fort que le Ré, ce qui obligera le violoniste à compenser avec l'archet cette différence d'intensité, et lui fera trouver l'instrument inégal. À propos des modes propres du violon, Savart observa vers 1830, par des mesures utilisant la méthode de Chladni[7], que le fond d'un bon violon et sa table d'harmonie ont leur premières fréquences propres écartées d'un demi-ton, évitant ainsi d'accentuer de manière trop forte l'une ou l'autre de ces fréquences (ces mesures furent faites sur des Stradivarius et Guarnierius désossés ! [20]).

[7] qui consiste à saupoudrer une plaque horizontale de fins granules et observer la disposition des grains lorsqu'on excite la plaque avec une vibration mécanique de fréquence donnée f. Aux fréquences propres de la plaque, les grains se disposent naturellement le long des lignes de nœuds du mode propre associé.

Dans tous les cas, en plus de sa fonction d'amplificateur, le résonateur agit directement sur la répartition spectrale du signal qu'il reçoit, et joue donc un rôle fondamental dans le timbre que produira l'instrument.

2.5.2 Enveloppe du son

Une troisième caractéristique fondamentale d'un son est son *enveloppe*, qui est en quelque sorte son « emballage», et dont la figure 1.20 nous avait donné un bel exemple. Celle-ci définit la manière dont apparaît, vit et disparaît le son musical. Le début du son, appelé *transitoire d'attaque*, est très riche en fréquences et joue un rôle crucial dans la reconnaissance de l'instrument. Privé de cette partie, de nombreux sons musicaux deviennent méconnaissables, en particulier s'il s'agit d'instruments à percussion. Dans la synthèse de sons musicaux, le transitoire d'attaque peut en partie être obtenu par une montée très rapide de l'intensité sonore au début de l'enveloppe, mais ce n'est pas toujours suffisant.

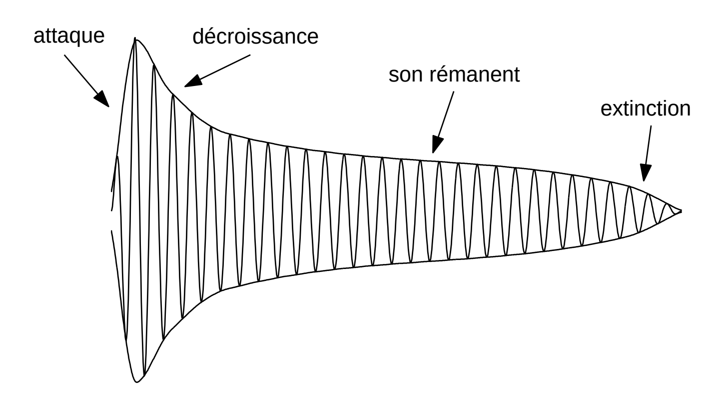


FIG. 2.18 – signal sinusoïdal $\sin 60\pi t$ modulé en amplitude par l'enveloppe $e(t)$.

Dans l'exemple suivant :

$$s(t) = e(t)\sin(60\pi t), \quad 0 \leq t \leq 1, \tag{2.23}$$

le signal sinusoïdal $\sin(60\pi t)$ est modulé en amplitude par une fonction $e(t) \geq 0$, l'enveloppe de $s(t)$, représentée sur la figure 2.18. L'enveloppe représentée est typique des instruments à percussion. Elle comporte quatre périodes principales : une période pendant laquelle le signal croît rapidement, l'attaque, d'une durée de quelques millisecondes à quelques centièmes de seconde, une période de décroissance suivie d'une autre période de décroissance plus lente, et une dernière période d'extinction du son. Ceci n'est bien sûr qu'un exemple, et chaque période peut elle-même être re-découpée en plusieurs parties. Pour les instruments à son entretenu, l'enveloppe peut avoir une allure très différente : l'attaque est souvent plus lente, et l'intensité peut ensuite être constante - voire croissante - pendant la majeure partie de la durée de la note.

Calcul de l'enveloppe

Pour calculer l'enveloppe, en particulier si l'on veut faire l'analyse du son d'un instrument, on peut s'inspirer de la technique de *démodulation d'amplitude* utilisée en communication radio. Pour cela, on fait d'abord passer le signal $s(t)$ dans un redresseur double alternance (un pont de diodes) qui délivre le signal $r(t) = |s(t)|$ (fig. 2.19). Ensuite, on fait passer le signal redressé dans un filtre passe-bas de réponse impulsionnelle $h(t)$, qui « lisse» les variations rapides pour ne conserver que l'enveloppe, et l'on obtient

$$e(t) \simeq a(h * |s|)(t),$$

où a est une constante qui dépend de s. En fait, on peut montrer le résultat suivant.

Soit un signal de la forme $s(t) = e(t)v(t)$, où $e(t) \geq 0$ pour tout t et $v(t)$ est T-périodique de fondamental $f_1 = 1/T$. Si l'on a $\widehat{e}(f) = 0$ pour tout $|f| > B$ avec $0 \leq B < f_1/2$, et si $h(t)$ est le filtre passe-bas idéal (1.26) de fréquence de coupure $f_1/2$, alors on a

$$e(t) = \frac{1}{c_0}(h * |s|)(t), \quad c_0 = \frac{1}{T}\int_0^T |v(t)|\, dt.$$

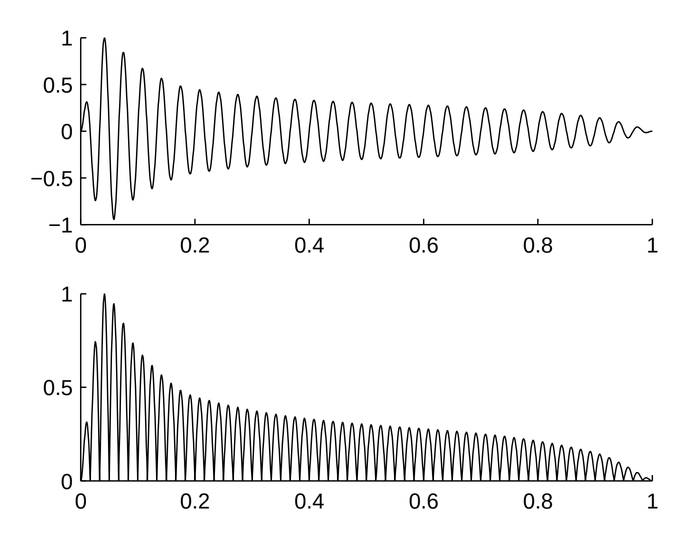


FIG. 2.19 – son d'origine $s(t)$ et son redressé $v(t) = |s(t)|$.

Même si les conditions portant sur $\widehat{e}(f)$ et B ne sont pas remplies, mais que $e(t)$ varie suffisamment lentement par rapport à $v(t)$, ce qui est le cas pour les sons instrumentaux, alors la différence entre $e(t)$ et $(h * |s|)(t)/c_0$ sera faible (cf. fig. 2.20). Nous retrouverons exactement les mêmes conditions dans le théorème de Shannon (cf. chapitre 5).

Utilisation de plusieurs enveloppes

Pour certains instruments, une seule enveloppe pour l'ensemble du son se révèle être insuffisante pour le décrire fidèlement. Il faut alors recourir à une enveloppe distincte $e_n(t)$

pour chaque harmonique et le son est alors représenté sous la forme

$$s(t) = \sum_n e_n(t) \sin(2\pi f_n t + \theta_n).$$

Pour analyser chaque enveloppe, on applique la technique du paragraphe précédent à chaque composante $e_n(t) \sin(2\pi f_n t + \theta_n)$, que l'on peut isoler par un filtre passe-bande adéquat. La figure 2.20 montre l'enveloppe des trois premiers harmoniques d'une note de piano, calculée selon la méthode du paragraphe précédent, ainsi que celles des harmoniques isolés 1 et 3, obtenus tous par filtrages passe-bande de la note. On constate en particulier que le troisième harmonique subit une chute d'intensité bien plus brutale que le premier.

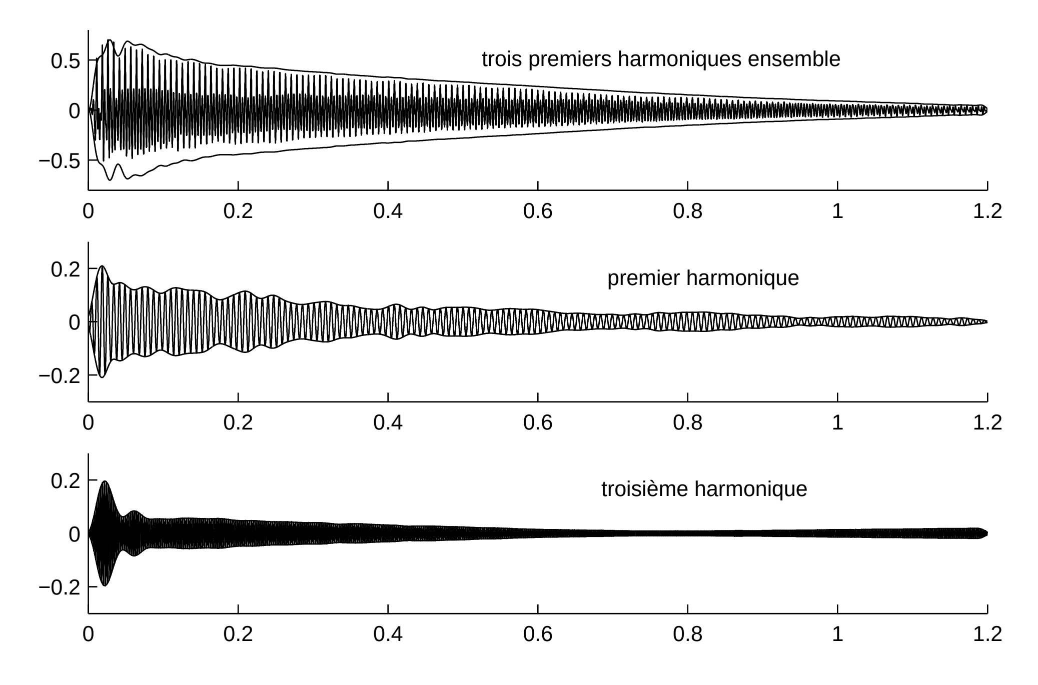


FIG. 2.20 – différentes enveloppes harmoniques du Do-2 (130.8Hz) d'un piano.

Terminons ce chapitre par deux spectrogrammes qui illustrent deux comportements différents des enveloppes des harmoniques (fig. 2.21 et 2.22). On y constate que dans un son de trompette, les harmoniques aigus arrivent en retard par rapport au fondamental (phénomène mis en évidence par les travaux de J.C. Risset, cf. [20]), ce qui constitue en partie la signature acoustique de cet instrument, alors que dans un son de piano, les harmoniques aigus s'éteignent plus vite que les graves.

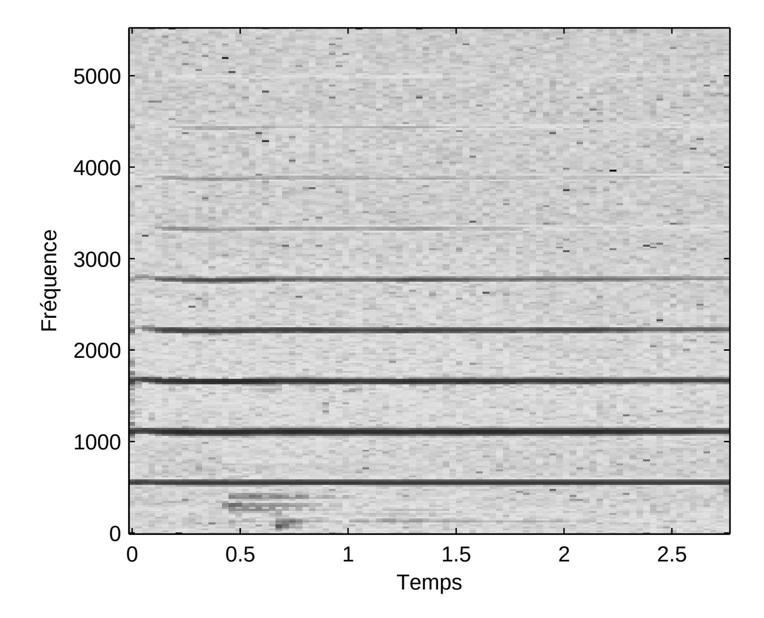


FIG. 2.21 – un extrait du trompettiste Miles Davis dans « The Sorcerer». Il s'agit d'une note tenue (Do♯-4), avec des harmoniques qui arrivent d'autant plus tard qu'ils sont aigus.

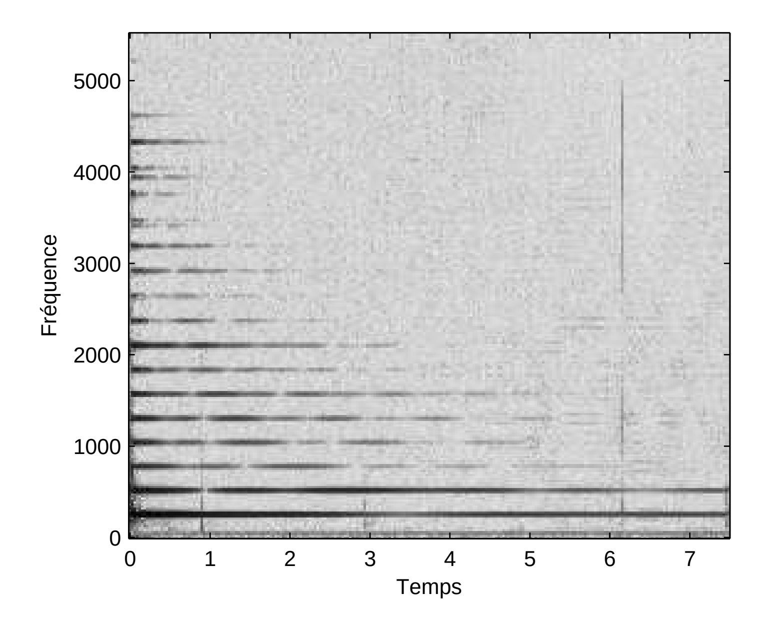


FIG. 2.22 – Do-3 d'un piano. Les harmoniques démarrent tous en même temps, mais leur durée décroît avec leur hauteur. On remarquera les battements produits par un léger désaccord entre les trois cordes.

2.6 Sujets d'étude

2.6.1 Vibration d'une corde (cas général) (**)

Le mouvement d'une corde libre de longueur L, fixée à ses deux extrémités, peut être déterminé par l'analyse de Fourier vue en cours. Il peut aussi être déterminé directement de la manière suivante. Nous savons déjà que ce mouvement est de la forme

$$u(x,t) = f(x-ct) + g(x+ct).$$

1) Montrez que la condition $u(0,t) = 0$ pour tout t entraîne que $g(y) = -f(-y)$ et donc

$$u(x,t) = f(x-ct) - f(-x-ct).$$

2) Montrez que la condition $u(L,t) = 0$ pour tout t entraîne que f est $2L$-périodique.
3) On écrit f sous la forme $f(x) = p(x) + q(x)$ où p et q sont aussi $2L$-périodiques, p étant paire ($p(x) = p(-x)$) et q impaire ($q(-x) = -q(x)$). On a ainsi

$$u(x,t) = p(x-ct) - p(x+ct) + q(x-ct) + q(x+ct).$$

Les conditions initiales sont données par

$$u(x,0) = u_0(x), \quad \partial_t u(x,0) = v_0(x).$$

Montrez que l'on a

$$q(x) = \frac{1}{2}u_0(x), \quad p(x) = -\frac{1}{2c}V_0(x) + A$$

où $V_0'(x) = v_0(x)$ et A est une constante.
4) Ces égalités ont lieu pour tout x si l'on suppose que u_0 et V_0 sont prolongées respectivement en des fonctions impaire et paire, $2L$-périodiques. Déduisez-en que

$$u(x,t) = \frac{1}{2}\left[u_0(x+ct) + u_0(x-ct)\right] + \frac{1}{2c}\left[V_0(x+ct) - V_0(x-ct)\right]$$

et que cette fonction est T-périodique en temps, avec $T = 2L/c$.

2.6.2 Corde pincée (*)

Pour une corde *pincée* (guitare, clavecin), les conditions initiales sont typiquement une vitesse nulle, et une position $u_0(x)$ affine par morceaux, faisant un angle à l'endroit où la corde est pincée.
1) En utilisant le sujet d'étude 2.6.1, représentez graphiquement la position $u(x,t)$ de la corde à quelques instants consécutifs, pris à l'intérieur d'une même période.
2) En utilisant l'analyse de Fourier, que peut-on dire à propos de l'harmonique de rang n si la corde est pincée exactement au point d'abscisse L/n ?

2.6.3 Corde frottée par un archet (*)

En 1877, étudiant le mouvement d'une corde *frottée* par un archet, Helmholtz observa que celle-ci se déformait d'une façon très particulière, représentée sur la figure 2.23. L'archet joue un double rôle : d'une part, il est responsable de la forme de cette onde, et d'autre part, son action permanente fournit l'énergie qui sera dépensée par la table d'harmonie. Nous allons retrouver ce mouvement de la corde en supposant que celle-ci est libre et sans amortissement, les conditions à l'instant $t = 0$ étant pour $0 < x < L$ et α quelconque :

$$u(x, 0) = 0, \quad \partial_t u(x, 0) = \alpha(L - x).$$

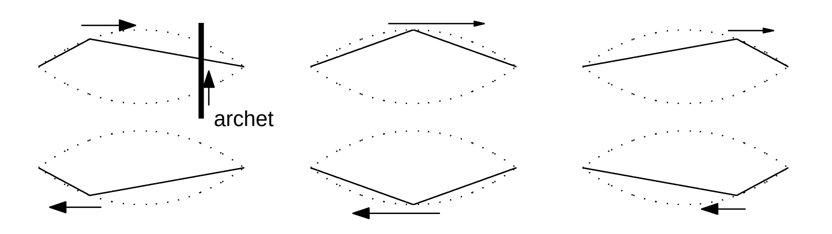


FIG. 2.23 – mouvement d'une corde frottée (à lire en tournant dans le sens des aiguilles d'une montre).

En utilisant le sujet d'étude 2.6.1, qui nous dit que u est T-périodique en temps avec $T = 2L/c$, montrez que sur la période $-T/2 \leq t \leq T/2$, on a

$$u(x, t) = \frac{\alpha}{4c}\left[(L - |x - ct|)^2 - (L - |x + ct|)^2\right].$$

Remarque : malgré les apparences, cette fonction est affine (par morceaux) par rapport à x et t, les termes de degré 2 s'annulant, et décrit bien le mouvement représenté sur la figure 2.23.

2.6.4 Corde réduite à un degré de liberté (**)

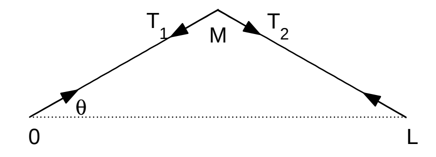


FIG. 2.24 – corde réduite à un point $M(t)$ de masse m.

Le modèle le plus simplifié d'une corde fixée à ses deux extrémités (cf. fig. 2.24) consiste à réduire celle-ci à un point mobile M de masse m, relié aux points d'attache de la corde

par deux fils élastiques sans masse, exerçant chacun sur le point M une force T_i qui a pour module la tension de la corde T :

$$T_1 = -T(\cos\theta, \sin\theta), \quad T_2 = T(\cos\theta, -\sin\theta).$$

La corde a pour longueur L, et le point M de coordonnées $(L/2, u(t))$ est astreint à se déplacer sur une ligne verticale (on ne considère pas les vibrations longitudinales).
1) En faisant l'approximation $\sin\theta \simeq \operatorname{tg}\theta$, montrez que la force F exercée sur le point M est

$$F = (0, -4uT/L).$$

2) En utilisant la relation fondamentale de la dynamique $F = m\gamma$, montrez qu'en l'absence de forces extérieures, on a

$$mu''(t) = -Ku(t)$$

où $K = 4T/L$ est la raideur du ressort vertical équivalent aux deux fils élastiques.
3) Montrez que les solutions sont des vibrations sinusoïdales de fréquence

$$f = \frac{1}{\pi L}\sqrt{\frac{T}{\mu}}$$

où $\mu = m/L$. On remarque que cette fréquence est inférieure au fondamental de la « vraie» corde, pour laquelle π est remplacé par 2. Cela s'explique par le fait que dans ce modèle simplifié, toute la masse a été concentrée au milieu, augmentant l'inertie de la corde.

2.6.5 Système couplé cordes-chevalet et phénomène de rémanence (***)

Chaque note de piano (hormis les graves) est composée de deux à trois cordes accordées à l'unisson. Typiquement, le son produit comporte deux phases : une première phase de décroissance rapide, suivie d'une phase à décroissance plus lente, que l'on appelle le *son rémanent*. Une première explication est la suivante : au début, ce sont les vibrations perpendiculaires à la table d'harmonie qui prédominent. Celles-ci s'amortissant rapidement, les vibrations parallèles à la table, d'amortissement plus lent, prennent alors le relais. Le son rémanent peut être également produit par un léger désaccord des cordes, comme l'illustre la figure 2.25. Celle-ci a été obtenue en fait par un modèle numérique simplifié que nous décrivons maintenant en détail.

Nous reprenons le modèle précédent, avec cette fois-ci plusieurs cordes reliées à un chevalet lui-même mobile. Chaque corde est réduite à un point M_i de masse m et de coordonnées $(L/2, u_i(t))$. Le chevalet est également réduit à un point Q de masse m_c et de coordonnées $(L, g(t))$. On suppose qu'il est relié à un point fixe par un amortisseur de raideur K_c et de résistance R. En présence d'une force extérieure $b(t)$ (qui sera dans la suite celle exercée par les cordes), l'équation qui régit le mouvement du chevalet est donc

$$m_c g''(t) = -Rg'(t) - K_c g(t) + b(t).$$

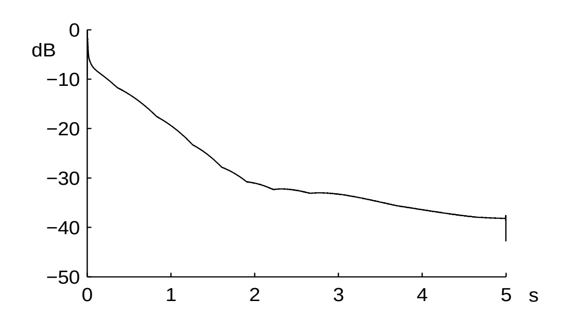


FIG. 2.25 – niveau sonore d'un son de piano. Le son rémanent provient ici d'un léger désaccord entre les cordes.

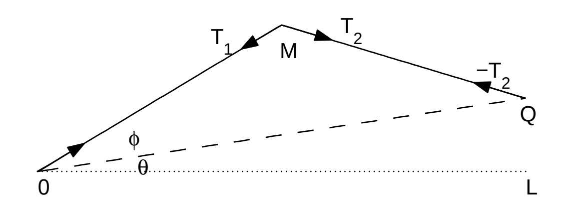


FIG. 2.26 – corde réduite à un point $M(t)$ de masse m, reliée au chevalet, lui-même réduit à un point $Q(t)$ de masse m_c.

1) On commence par examiner l'interaction d'une seule corde avec le chevalet (cf. fig. 2.26). Les forces s'exerçant sur le point M sont

$$T_1 = -T(\cos(\phi+\theta), \sin(\phi+\theta)), \quad T_2 = T(\cos(\phi-\theta), -\sin(\phi-\theta))$$

où T est la tension de la corde. En utilisant les approximations $\sin\theta \simeq \operatorname{tg}\theta, \cos\theta \simeq 1, \ldots$, montrez que la composante verticale des forces exercées sur M vaut $-2T(2u-g)/L$ et déduisez-en que

$$mu''(t) = -Ku(t) + Kg(t)/2$$

avec $K = 4T/L$. Montrez de même que la composante verticale des forces exercées par la corde sur Q vaut $2T(u-g)/L$ et déduisez-en que

$$m_c g''(t) = -Rg'(t) - K_c g(t) - Kg(t)/2 + Ku(t)/2.$$

2) On considère maintenant le cas de deux ou trois cordes $u_i(t)$ soumises chacune à une tension T_i, et l'on pose :

$$K_i = \frac{4T_i}{L}, \quad K_s = \sum_{i=1}^{2 \text{ ou } 3} K_i.$$

Montrez que les équations du système complet deviennent

$$\begin{cases} mu_i''(t) = -K_i u(t) + K_i g(t)/2, \quad i = 1, 2, \ldots \\ m_c g''(t) = -Rg'(t) - (K_c + K_s/2)g(t) + \sum K_i u_i(t)/2. \end{cases} \tag{2.24}$$

3) Pour la résolution numérique (on traite ici le cas de 2 cordes), on pose

$$u_i' = v_i, \quad g' = h,$$
$$X = (u_1, u_2, g, v_1, v_2, h).$$

Montrez que le système différentiel (2.24) peut s'écrire sous la forme matricielle

$$X'(t) = AX(t)$$

avec

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -K_1/m & 0 & K_1/(2m) & 0 & 0 & 0 \\ 0 & -K_2/m & K_2/(2m) & 0 & 0 & 0 \\ K_1/(2m_c) & K_2/(2m_c) & -(K_c+K_s/2)/m_c & 0 & 0 & -R/m_c \end{bmatrix}.$$

La solution est alors

$$X(t) = \exp(tA)X_0$$

où le vecteur X_0 contient les conditions initiales $X(0)$. La résolution numérique consiste par exemple à choisir un pas de temps τ, à calculer $M = \exp(\tau A)$, et $X_n = X(n\tau)$ s'obtient par récurrence :

$$X_{n+1} = MX_n.$$

Les paramètres utilisés pour la figure 2.25 ont été les suivants :

$$m = 1, \quad K_i = (2\pi f_i)^2 m \text{ avec } f_1 = 500\text{Hz et } f_2 = 500.1\text{Hz},$$
$$m_c = 300, \quad K_c = (2\pi f_c)^2 m_c \text{ avec } f_c = 50\text{Hz},$$
$$R = 10\sqrt{K_c m_c},$$
$$X_0 = (0, 0, 0, 1, .9, 0).$$

2.6.6 Calcul de l'inharmonicité pour une corde réelle (***)

L'équation qui régit le mouvement d'une corde réelle, possédant un diamètre non nul et donc une certaine raideur, est une combinaison de celles d'une corde idéale et d'une tige. Son expression est la suivante :

$$\partial_{t^2} u(x,t) - c^2 \partial_{x^2} u(x,t) + \lambda c^2 \partial_{x^4} u(x,t) = 0$$

où $c^2 = T/\mu$, T étant la tension de la corde et μ sa masse linéique, et $\lambda = \pi r^4 E/(4T)$, E étant le module de Young de la corde et r son rayon. Nous étudions les solutions harmoniques de la forme $u(x,t) = \varphi(x)\exp(2i\pi f t)$, qui n'existent que pour certaines valeurs f_n de la fréquence, que nous allons déterminer. Nous verrons que ces fréquences ne sont plus des harmoniques comme dans le cas d'une corde idéale, mais des partiels.

1) Montrez que φ est solution de l'équation

$$\lambda \varphi^{(4)}(x) - \varphi''(x) - k^2 \varphi(x) = 0 \tag{2.25}$$

avec $k = 2\pi f/c$.
2) Les solutions de cette équation sont des combinaisons linéaires d'exponentielles de la forme $\exp(Kx)$. Montrez que $\exp(Kx)$ est une solution si et seulement si

$$\lambda K^4 - K^2 - k^2 = 0,$$

et déduisez-en que la solution générale de (2.25) est de la forme

$$\varphi(x) = \alpha \exp(K_1 x) + \beta \exp(-K_1 x) + \gamma \exp(iK_2 x) + \delta \exp(-iK_2 x) \tag{2.26}$$

avec

$$K_1^2 = \frac{1 + \sqrt{1 + 4\lambda k^2}}{2\lambda}, \quad K_2^2 = \frac{-1 + \sqrt{1 + 4\lambda k^2}}{2\lambda}.$$

3) Les 4 constantes dans (2.26) doivent obéir aux conditions imposées aux extrémités de la corde. Nous supposons que celles-ci sont les suivantes :

$$\varphi(0) = 0, \quad \varphi''(0) = 0,$$
$$\varphi(L) = 0, \quad \varphi''(L) = 0.$$

Montrez que les deux premières conduisent à $\alpha = -\beta$ et $\gamma = -\delta$. La solution est donc de la forme

$$\varphi(x)/2 = \alpha \operatorname{sh}(K_1 x) + i\gamma \sin(K_2 x).$$

Montrez alors que les deux dernières conditions conduisent à des solutions non nulles si et seulement si

$$\sin K_2 L = 0$$

(on calculera le déterminant du système homogène de 2 équations à deux inconnues α et γ).
4) On a donc la condition suivante qui porte sur $K_2 = K_2(n)$, la dépendance par rapport à n portant en fait sur $k = k_n = 2\pi f_n/c$:

$$K_2 L = n\pi.$$

En posant $B = \lambda\pi^2/L^2$, déduisez de cette condition que

$$k_n^2 = \frac{n^2\pi^2}{L^2}(1 + Bn^2),$$

et montrez finalement que le partiel f_n a pour expression

$$f_n = n f_1^0 \sqrt{1 + Bn^2}$$

où $f_1^0 = c/(2L)$ est le fondamental de la corde en l'absence d'inharmonicité. On remarque que pour $r = 0$ (la corde idéale), on a $\lambda = 0$ et l'on retrouve la formule $f_n = nf_1$.

2.6.7 Fréquence de coïncidence d'une onde dans une plaque (***)

Le résonateur de nombreux instruments de musique est constitué d'une plaque en bois. Celui-ci (épicéa par exemple) est choisi pour que la vitesse de propagation des ondes transversales y soit élevée. Nous allons voir en quoi cela est utile pour obtenir un bon rayonnement de l'onde sonore.

Pour simplifier, on considère une plaque infinie, dont la tranche médiane est située au repos dans le plan xOy, vibrant verticalement à la fréquence f, et dont la position sur l'axe des z est décrite par une fonction de la forme

$$u(x, y, t) = \cos(kx) \exp(i\omega t),$$

avec $\omega = 2\pi f$. L'équation des ondes transversales pour une plaque s'écrit

$$\partial_{t^2} u + \frac{h^2 c_L^2}{12} \Delta^2 u = 0, \quad c_L^2 = \frac{E}{\rho(1 - \nu^2)}$$

où h est l'épaisseur de la plaque, E, ρ et ν sont respectivement le module de Young, la densité et le coefficient de Poisson du matériau supposé homogène et isotrope, c_L est la vitesse des ondes longitudinales dans la plaque, et $\Delta^2 u := \Delta(\Delta u)$, qui se réduit ici à $\Delta^2 u = \partial_{x^4} u$.

1) En supposant $k > 0$, montrez que

$$k = \sqrt{\frac{\omega\sqrt{12}}{hc_L}},$$

et que l'onde transversale stationnaire $u(x, y, t)$ est la superposition de deux ondes progressives se propageant dans la plaque en sens inverse à la vitesse (qui dépend de la fréquence)

$$c(\omega) = \frac{\omega}{k} = \sqrt{\frac{\omega h c_L}{\sqrt{12}}}.$$

2) Les vibrations de la plaque engendrent une onde acoustique dans l'air occupant le demi-espace $z \geq 0$. Les composantes du vecteur vitesse de l'air sont supposées nulles sauf la composante en z notée $v_z(x, y, z, t)$, qui coïncide en $z = 0$ avec la vitesse de la plaque :

$$v_z(x, y, 0, t) = i\omega \cos(kx) \exp(i\omega t).$$

On suppose que $v_z(x, y, z, t)$ est de la forme

$$v_z(x, y, z, t) = i\omega \cos(kx) \exp(i\omega t - i\kappa z). \tag{2.27}$$

Sachant que v_z est solution de l'équation des ondes (1.9), montrez que

$$\kappa^2 = \omega^2 \left(\frac{1}{c^2} - \frac{1}{c(\omega)^2} \right).$$

Une première conclusion est que si $c(\omega) < c$, alors κ est imaginaire pur, et l'onde acoustique aura une décroissance exponentielle en $\exp(-\alpha z)$ avec $0 < \alpha = i\kappa$ (pour des raisons physiques,

la solution croissant exponentiellement est écartée). De plus, étant stationnaire, son intensité sur une période T est nulle (cf. § 1.6.4). Par contre, si $c(\omega) > c$, alors κ est réel, l'onde est progressive, le comportement en z est sinusoïdal et le son « porte». La fréquence f à laquelle se produit l'égalité $c(\omega) = c$ est appelée *fréquence de coïncidence.*
3) On suppose maintenant que $c(\omega) > c$, et l'on s'intéresse à l'intensité sonore en un point x, y, z. La vitesse réelle, toujours notée v_z, est (prendre la partie réelle de (2.27)) :

$$v_z(x, y, z, t) = -\omega \cos(kx) \sin(\omega t - \kappa z).$$

En utilisant l'équation d'état $\partial_t p_a = -c^2 \rho_0 (\partial_x v_z + \partial_y v_z + \partial_z v_z)$, montrez que la pression acoustique p_a est

$$p_a(x, y, z, t) = c^2 \rho_0 [k \sin(kx) \cos(\omega t - \kappa z) - \kappa \cos(kx) \sin(\omega t - \kappa z)].$$

Déduisez-en que l'intensité au point x, y, z a pour valeur

$$I = \frac{c\rho_0 \omega^2 \cos^2(kx)}{2} \sqrt{1 - \frac{c^2}{c(\omega)^2}}.$$

Conclusion ?

2.6.8 Résonance du bourdon (**)

Le bourdon de l'orgue peut être considéré de façon simplifiée comme un tube commandé en pression à l'entrée en $x = 0$, la particularité de ce tube étant d'être fermé à l'autre extrémité en $x = L$ (avec les notations du cours), où la vitesse s'annulle donc :

$$v(L, t) = 0 \text{ pour tout } t.$$

1) En utilisant la relation d'Euler, montrez que la pression $p(x, t)$ à l'intérieur du tube vérifie la condition aux limites

$$\partial_x p(L, t) = 0.$$

2) Reprenez alors l'étude de la résonance du tube commandé en pression du §2.4.1 avec maintenant

$$\begin{cases} \partial_{t^2} p(x, t) - c^2 \partial_{x^2} p(x, t) = 0, & \text{dans le tube,} \\ p(0, t) = p_0(t), & \text{à l'entrée,} \\ \partial_x p(L, t) = 0, & \text{à l'autre extrémité,} \end{cases}$$

et montrez que les fréquences propres sont les mêmes que pour le tube ouvert commandé en vitesse :

$$f_n = \frac{(n - 1/2)c}{2L}, \quad n \geq 1.$$

2.6.9 Résonance d'un tube cylindrique avec commande mixte (**)

Nous étudions le cas d'une onde plane harmonique $p(x,t) = \varphi(x)\exp(2i\pi ft)$ dans un tube cylindrique de longueur L, engendrée par une *commande mixte*, c'est-à-dire que les conditions aux limites s'écrivent pour a et b réels (c'est un choix particulier, il y en a d'autres) :

$$\begin{cases} ap(0,t) + b\partial_x p(0,t) = \exp(2i\pi ft), & \text{à l'entrée,} \\ p(L,t) = 0, & \text{à la sortie.} \end{cases}$$

1) On rappelle que $\varphi(x)$ est de la forme

$$\varphi(x) = \alpha\exp(ikx) + \beta\exp(-ikx),$$

où $k = 2\pi f/c$. Montrez que les conditions aux limites ci-dessus imposent

$$\begin{cases} \alpha(a+ikb) + \beta(a-ikb) = 1, \\ \alpha\exp(ikL) + \beta\exp(-ikL) = 0, \end{cases}$$

et déduisez-en que

$$\alpha = \frac{\exp(-ikL)}{z\exp(-ikL) - \overline{z}\exp(ikL)}, \quad \beta = \frac{-\exp(ikL)}{z\exp(-ikL) - \overline{z}\exp(ikL)},$$

où $z = a + ikb$.
2) En écrivant z sous la forme $z = r\exp(i\theta)$, montrez que les fréquences de résonance, c'est-à-dire les valeurs de f pour lesquelles α et β ne sont pas définis, sont les nombres

$$f_n = \frac{(n\pi + \theta)c}{2\pi L}, \quad n \in \mathbb{Z}.$$

Que retrouve-t-on pour les cas particuliers $(a,b) = (1,0)$ et $(a,b) = (0,1)$?

2.6.10 Résonance d'un tube conique (1) (**)

Une différence majeure entre la clarinette d'une part, le saxophone, le hautbois et le basson d'autre part, est que dans le premier cas la perce est cylindrique, alors que dans le second elle est conique. Là réside en grande partie la différence de timbre entre ces instruments, qui sont par ailleurs tous les trois à anche (simple ou double), et donc fonctionnant en première approximation sous un régime de commande en vitesse. Alors que les harmoniques pairs sont quasiment absents d'un son de clarinette, nous allons voir que ce n'est plus du tout le cas du hautbois et du saxophone.

On considère un tube conique tronqué (dont le prolongement aurait son sommet à l'origine), délimité par les extrémités $r = a$ et $r = b$, avec $0 < a < b$, la longueur du tube étant $L = b - a$ (on utilise les coordonnées sphériques avec $r = \|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2}$). En régime harmonique, la pression acoustique à l'intérieur du tube est de la forme

$$p(r,t) = \left[\alpha\frac{\exp(ikr)}{r} + \beta\frac{\exp(-ikr)}{r}\right]\exp(2i\pi ft).$$

1) On note $\mathbf{n}$ le vecteur unitaire sortant normal à la surface délimitant l'entrée du tube ($r = a$). Sachant que l'équation d'Euler pour une onde de pression acoustique s'écrit

$$-\nabla p(r,t) = \rho_0 \partial_t v(r,t),$$

montrez que la commande en vitesse $v(r,t).\mathbf{n} = \exp(2i\pi ft)$ imposée à l'entrée $r = a$ devient

$$\partial_r p(a,t) = 2i\pi\rho_0 f \exp(2i\pi ft).$$

2) La pression étant supposée nulle à la sortie du tube, montrez que les conditions aux limites donnent respectivement en $r = a$ et $r = b$

$$\begin{cases} \alpha z \exp(ika) + \beta \overline{z} \exp(-ika) = 2i\pi a^2 \rho_0 f, \\ \alpha \exp(ikb) + \beta \exp(-ikb) = 0, \end{cases}$$

où $z = ika - 1$, et déduisez-en que

$$\alpha = \frac{2i\pi a^2 \rho_0 f \exp(-ikb)}{z \exp(-ikL) - \overline{z} \exp(ikL)}, \quad \beta = -\alpha \exp(2ikb).$$

3) Montrez que les fréquences de résonance, c'est-à-dire les valeurs de f pour lesquelles α et β ne sont pas définis, sont solution de l'équation transcendante

$$\mathrm{tg}\,(2\pi fL/c) + 2\pi fa/c = 0,$$

et déduisez-en que les fréquences de résonance sont de la forme

$$f_n = \frac{c\,(-\mathrm{arctg}\,(2\pi f_n a/c) + n\pi)}{2\pi L}, \quad n \in \mathbb{Z}.$$

4) Tracez graphiquement les courbes d'équation $y = \mathrm{tg}\,(2\pi fL/c)$ et $y = -2\pi fa/c$, dont les points d'intersection donnent en abscisse les valeurs f_n, et déduisez-en que pour $a \ll L$ et n pas trop grand, on a

$$f_n \simeq \frac{nc}{2L}.$$

Conclusion ?

2.6.11 Résonance d'un tube conique (2) (**)

Reprendre le sujet d'étude 2.6.10, en supposant cette fois-ci que l'on a une commande en pression, c'est-à-dire que les conditions imposées aux extrémités sont $p(a,t) = \exp(2i\pi ft)$ et $p(b,t) = 0$. Solution : les fréquences de résonance sont exactement les nombres

$$f_n = \frac{nc}{2L}.$$

On constate ainsi que la différence provenant du type de commande pour un tube cylindrique, devient négligeable pour un tube conique (pour ce qui concerne les fréquences de résonance).

2.7 Travaux pratiques sur ordinateur

Cette seconde série de travaux pratiques est consacrée à la fabrication d'un son synthétique imitant un instrument de musique acoustique, ainsi qu'à quelques expériences sur le timbre. La dernière partie (facultative) ne peut être faite que si vous avez traité le sujet d'étude 2.6.5.

2.7.1 Créez votre synthétiseur

Nous allons écrire une fonction MATLAB chargée de créer des sons par *synthèse addiditive* (addition de sinus, cf. aussi chapitre 6) et utilisation d'une *enveloppe* pour contrôler l'intensité au cours du temps. La synthèse additive fera appel à la fonction `synthad.m` déjà utilisée aux TP du chapitre 1.

Écrivez votre fonction instrument

Créez une fonction ayant pour en-tête

$$\texttt{function s} = \texttt{instrument(f1,T,Fe)}$$

Rappelons que le fichier s'appelle alors `instrument.m` (vous pouvez remplacer le mot 'instrument' par celui de votre choix). Créons d'abord un son de la forme

$$s(t) = \sum_{k=1}^{np} \alpha_k \sin(2\pi f_1 h_k t), \quad 0 \leq t \leq T.$$

Comme il s'agit là d'un instrument particulier qui est programmé, le nombre d'harmoniques ou de partiels np, leur amplitude α_k et leur fréquence « normalisée» h_k, sont tous définis *par vous* dans la fonction elle-même : la fréquence du partiel numéro k est alors $f_1 h_k$. En principe (c'est l'effet de la « normalisation»), on a $h_1 = 1$ ($h_1 = 0.5$ pour une cloche), et $f_1 h_1 = f_1$ ($f_1/2$ pour une cloche) est la fréquence du premier harmonique ou partiel. Une fois définies ces valeurs, vous pouvez faire appel, à l'intérieur même de la fonction `instrument`, à la fonction `synthad`.

Quelques suggestions pour le choix de la répartition des harmoniques ou partiels : l'exemple de la cloche du chapitre 1, les deux types de lames, les deux types de tubes ou la membrane de ce chapitre. Il ne vous reste plus qu'à tester votre fonction, par exemple en exécutant les commandes suivantes :

```
Fe = 22050 ;
s = instrument(220,3,Fe) ;
sound(s,Fe) ;
```

Ajoutez une enveloppe

Télé-chargez la fonction `envelop.m` sur ma page web. Pour comprendre ce que fait cette fonction, exécutez les commandes suivantes :

```
t = [0 .1 .4 1] ; a = [0 1 .3 0] ;
env = envelop(t,a,Fe) ;
plot(env) ;
```

puis recommencez en faisant varier les valeurs (et leur nombre) dans les vecteurs `t` et `a`. Vous y êtes ? Pour ajouter une enveloppe au son crée par votre fonction `instrument`, il vous suffit alors d'y rajouter à la fin (du fichier) les lignes de commande suivantes *après avoir spécifié vous-même* dans la fonction les valeurs numériques des vecteurs `t` et `a`, en veillant à ce que le premier élément dans `t` soit 0 et le dernier la durée `T` :

```
env = envelop(t,a,Fe) ;
s = env.*s ;
```

Testez à nouveau votre fonction.

Et jouez de votre instrument

Télé-chargez le fichier `play.m` (toujours au même endroit), éditez-le et lisez-le. Vous allez y apporter quelques modifications. Ce programme calcule les 13 notes d'une gamme chromatique en partant d'une fréquence grave `f0` que vous pouvez choisir. Vous devez aussi préciser dans ce fichier l'instrument (le nom de la fonction) que vous voulez jouer. Ceci fait, lancez la commande `play` et jouez !

Remarque 2.3 *Sous le système* `windows`*, il existe une application nommée 'Vienna' qui permet de créer des 'soundfonts' (banque de sons) à partir d'échantillons, ce qui permet ensuite de les jouer, soit à partir d'un clavier virtuel à l'écran, soit à partir d'un vrai clavier connecté à l'ordinateur par une prise MIDI. Pour en savoir plus sur ce sujet, voir par exemple le site http ://perso.club-internet.fr/michbuze/Musique/midi.htm. De tels outils existent aussi sous le système* `linux`*, mais posent pour le moment des problèmes d'installation en réseau.*

2.7.2 Modifiez le timbre de votre instrument

Maintenant que votre « synthétiseur» fonctionne, vous pouvez vous amuser à modifier le timbre de votre instrument en jouant sur les poids des harmoniques ou partiels dans votre fichier `instrument.m`, et écouter l'effet produit *via* le programme `play`. Par exemple, que se passe-t-il si l'on supprime le premier harmonique ? Si l'on modifie la forme de l'enveloppe ? Ou le poids relatif des harmoniques ? Si l'on introduit de l'inharmonicité ?

2.7.3 Son rémanent

Les notations et les données sont celles du sujet d'étude 2.6.5. Créez un fichier de commandes intitulé `unisson.m` dans lequel vous allez programmer les calculs nécessaires. On décrit ici le modèle à deux cordes. Il s'agit essentiellement de programmer une boucle pour calculer $X_{n+1} = MX_n$ avec $X_n = X(n\tau)$ et $M = \exp(\tau A)$. Pour cela il faut initialiser les données, en particulier X_0 et A (on pourra utiliser celles mentionnées dans le sujet d'étude 2.6.5, mais vous pourrez trouver plus amusant de chercher par vous même les valeurs des paramètres qui produisent le phénomène de rémanence).

Initialisation de X_0 : il s'agit là d'un vecteur colonne, la commande s'écrit par exemple

```
X = [0 0 0 1 .9 0]' ;
```

(noter le *prime* qui a pour effet de transposer, ici changer une ligne en une colonne).

Initialisation de A : appuyez-vous sur l'exemple suivant : pour initialiser la matrice

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 2 & 2 & 0 \end{bmatrix},$$

la commande s'écrit

```
A = [1 2 3 ; 0 1 1 ; 2 2 0] ;
```

(notez le ; pour passer à la ligne suivante). Le calcul de M se fait par la commande

```
M = expm(tau*A) ;
```

où `tau` est le pas de temps choisi pour représenter le son.

Le son lui-même est donné par la position du chevalet, troisième composante du vecteur X. L'initialisation du vecteur son et la boucle de calcul s'écrivent, si `nt` est le nombre de pas de temps :

```
son = zeros(1,nt) ;
son(1) = X(3) ;
for n = 2 :nt
   X = M*X ;
   son(n) = X(3) ;
end
```

Enfin, pour écouter et observer graphiquement le son produit, les commandes sont :

```
soundsc(son,1/tau) ;
plot(20*log10(abs(son)+1e-3)) ;
```

Chapitre 3

Les gammes

Dans l'étude des corps vibrants du chapitre 2, nous avons constaté que ceux-ci produisent selon les cas des harmoniques (cordes et tubes) ou des partiels (cloches, membranes, percussions). Il semblerait que l'homme ait été davantage sensible aux premiers qu'aux seconds pour construire ses échelles musicales. À cela on peut avancer l'explication suivante : les cordes (provenant de l'arc du chasseur) et les tubes (premières flûtes en os) produisent des sons plus longs que les percussions, et il est bien plus facile de percevoir les harmoniques des premiers en fonction de leur degré de consonance (cf. § 2.1.1), que les partiels des seconds. Aussi est-il probable que la forte consonance de la quinte[1], correspondant à un rapport de fréquence 3/2, se soit imposée très tôt dans l'histoire musicale.

Nous avons vu que la division successive d'une corde de longueur L en segments de longueurs $L/1$, $L/2$, $L/3$, $L/4$, $L/5$... produisait la suite des harmoniques, de fréquences égales à 1, 2, 3, 4, 5... fois le fondamental f. Nous voyons apparaître là deux suites de nombres, la première harmonique, la seconde arithmétique. Remarquons qu'elles seraient apparues dans l'ordre inverse si au lieu de raccourcir la corde, nous avions multiplié sa longueur par 1, 2, 3, 4, 5 pour obtenir des fréquences égales à $f/1$, $f/2$, $f/3$, $f/4$, $f/5$. Les treize premiers harmoniques correspondent en gros aux notes données dans le tableau suivant, où la ligne supérieure indique aussi bien le rang de l'harmonique que le rapport de sa fréquence à celle du fondamental.

1	2	3	4	5	6	7	8	9	10	11	12	13
Do	Do	Sol	Do	Mi	Sol	Si♭-	Do	Ré+	Mi-	Fa♯-	Sol	Sol♯+

En écriture musicale, ce sont à quelques cents près (cf. § 2.5.1) les notes suivantes :

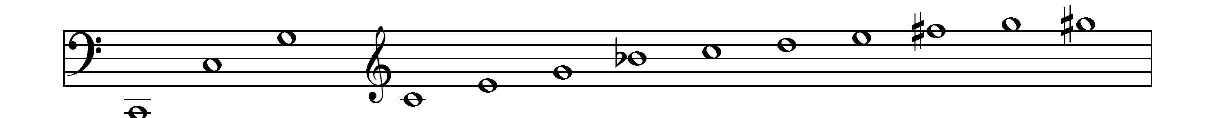


Fig. 3.1 – La suite des treize premiers harmoniques du Do-1.

[1]résultant de la coïncidence entre l'harmonique 3 de la note grave et l'harmonique 2 de la note aiguë.

Si nous ramenons toutes ces notes à l'intérieur d'une même octave en divisant les fréquences par une puissance de 2 convenable, et les ordonnons par fréquence croissante, nous obtenons les rapports de fréquence suivants :

Do	Ré	Mi-	Fa♯-	Sol	Sol♯+	Si♭-	Do
1	9/8	5/4	11/8	3/2	13/8	7/4	2

(3.1)

Ces notes ne correspondent à aucune gamme classique, mais on les entend très bien par exemple sur une flûte traversière lorsqu'on souffle très doucement, sans faire résonner le tube.

Ce sont essentiellement sur les deux premiers harmoniques distincts de l'octave, Sol et Mi, que porteront les innombrables discussions qui ont eu lieu au cours des siècles sur les intervalles de la gamme [2].

3.1 La gamme de Pythagore

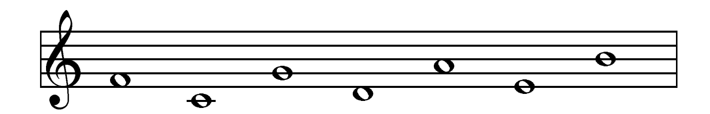


Fig. 3.2 – gamme diatonique de Pythagore. Toutes les quintes et quartes *entre notes consécutives* sont « justes», c'est-à-dire dans un rapport 3/2 pour les quintes, 4/3 pour les quartes (rapport des fréquences aiguë/grave).

Au VIème siècle av. J.-C., Pythagore privilégiait la simplicité arithmétique des rapports de longueurs. Mis à par le rapport d'octave 2, le rapport de quinte 3/2 est le plus « simple» qui apparaisse dans le tableau (3.1). Aussi Pythagore mit en avant la quinte pour construire, par reproduction répétée de cet intervalle, la gamme diatonique qui porte son nom. C'est le cycle bien connu des quintes Fa, Do, Sol, Ré, La, Mi, Si, qui donne les rapports de fréquence suivants :

Do	Ré	Mi	Fa	Sol	La	Si	Do
1	9/8	81/64	4/3	3/2	27/16	243/128	2

Cette gamme est arithmétiquement très « élégante», au sens qu'elle ne donne que deux types d'intervalle entre notes consécutives : le ton, de rapport 9/8, et le demi-ton, de rapport 256/243. Cependant, mise à part la quinte (et bien sûr l'octave), elle ne coïncide pas avec les harmoniques. Par exemple, la tierce de rapport $81/64 \simeq 1.2656$ est plus grande que celle de l'harmonique correspondant du tableau 3.1 qui vaut $5/4 = 1.25$. La différence, peu sensible à l'intérieur d'une *mélodie*, le deviendra par contre en *harmonie*, c'est-à-dire dans un accord, à cause du phénomène des battements (cf. § 2.1.1) : pour un Do de fréquence 262Hz, la tierce harmonique Do-Mi ne produira aucun battement, alors que la tierce pythagoricienne Do-Mi produira un battement de $|(5\times1-4\times81/64)|\times262\text{Hz} \simeq 16\text{Hz}$, parfaitement audible[2]. Pour nos oreilles habituées au tempérament égal que nous verrons plus loin, la tierce pythagoricienne

[2] à condition bien sûr que l'harmonique 5 du Do et l'harmonique 4 du Mi, entre qui se produit ce battement, soient présents dans les notes.

est perçue comme « brillante» ou « dure». Le battement de la tierce tempérée est nettement plus lent : environ 10Hz.

3.2 La gamme de Zarlino

Au IVème siècle av. J.-C., Aristoxène le Musicien, élève d'Aristote, recommandait dans ses *Éléments d'harmonique* de « ne pas tourner le dos à la sensation», et de se fier à l'oreille plutôt qu'à la raison mathématique. Aussi proposait-il de construire l'accord majeur sur la base des harmoniques, donc dans les rapports de fréquence 1, 5/4, 3/2. C'est un accord qui ne produit aucun battement et sonne très « pur». Il était jugé par Aristoxène et ses disciples plus esthétique et plus conforme à la physique.

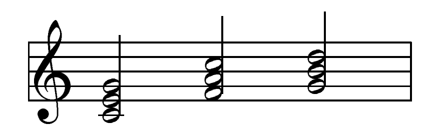


Fig. 3.3 – les trois accords majeurs servant à construire la gamme de Zarlino. Tierces et quintes harmoniques dans chaque accord. Les autres font ce qu'elles peuvent !

Au XVIème siècle, cette idée fut reprise par Zarlino, qui partit des trois accords majeurs Fa-La-Do, Do-Mi-Sol et Sol-Si-Ré, avec comme rapports de fréquence 1, 5/4, 3/2, pour construire sa gamme diatonique, appelée aussi gamme des physiciens. Les rapports de fréquence sont les suivants :

Do	Ré	Mi	Fa	Sol	La	Si	Do
1	9/8	5/4	4/3	3/2	5/3	15/8	2

Dans cette gamme, les trois accords majeurs de tonique Fa, Do et Sol sont donc très consonants et ne produisent aucun battement. Mais les choses se gâtent singulièrement quand on change de tonalité. Par exemple, la quinte Ré-La a un rapport de fréquence égal à $40/27 \simeq 1.48$, inférieur à la quinte « juste» de rapport 1.5. Pour un Ré de fréquence 294Hz, la quinte Ré-La de Zarlino produira un battement de $(3 \times 1 - 2 \times 40/27) \times 294\text{Hz} \simeq 11\text{Hz}$, ce qui est épouvantable pour une quinte !

Par ailleurs, il y a maintenant trois types d'intervalles entre notes consécutives (contre deux dans la gamme de Pythagore) : le demi-ton de rapport 16/15, et deux types de tons : les tons Do-Ré, Fa-Sol et La-Si de rapport 9/8, et les tons Ré-Mi et Sol-La de rapport 10/9 ! L'écart entre ces deux types correspond à un rapport de fréquence égal à $(9/8)/(10/9) = 81/80$, intervalle appelé *comma*, et valant approximativement un neuvième de ton.

Tout cela n'était pas très satisfaisant, et tout particulièrement dans le contexte de la musique occidentale où se développait la polyphonie combinée avec le désir de pouvoir transposer dans toutes les tonalités.

3.3 Les gammes tempérées

Le clavecin bien tempéré de J. S. Bach fut composé au début du XVIIIème siècle. C'est un ensemble de préludes et fugues écrits dans les vingt-quatre tonalités, majeures et mineures. Comme l'indique le titre de l'œuvre, jouer ces préludes et fugues nécessite donc une gamme « bien tempérée» (qui ne signifie d'ailleurs pas tempérament égal), et le système de Zarlino avec certaines quintes trop raccourcies est pour cela tout à fait inadéquat.

Qu'à cela ne tienne, revenons à Pythagore et poursuivons le cycle des quintes amorcé, jusqu'à avoir parcouru les douze demi-tons de la gamme. Notons au passage que la gamme chromatique ainsi obtenue était déjà connue des chinois au IVème siècle av. J.-C., qui s'en servaient pour transposer les gammes pentatoniques typiques de la musique orientale. En partant de Fa, le cycle Fa, Do, Sol, Ré, La, Mi, Si, Fa♯, Do♯, Sol♯, Ré♯, La♯, Mi♯ va donner comme rapport de fréquence entre la première et la dernière note $(3/2)^{12} \simeq 129.7463$. Par ailleurs, on aimerait que ce Mi♯ soit un Fa, ne serait-ce que pour limiter le nombre des notes dans un clavier. Or le Fa le plus proche correspond au rapport $2^7 = 128$. Tout le problème est là :

$$129.7463 \simeq (3/2)^{12} \neq 2^7 = 128,$$

ce qui rend impossible d'avoir *à la fois* des octaves justes et des quintes justes. Le rapport entre ces deux notes, $(3/2)^{12}/2^7 \simeq 1.0136$ s'appelle le *comma pythagoricien*. Il fallait donc trouver un compromis. Celui-ci fut trouvé par Werckmeister à la fin du XVIIème siècle. En posant l'équation 12 quintes = 7 octaves et en privilégiant la justesse des octaves, il décida de répartir le comma excédentaire entre les 12 quintes en les raccourcissant chacune légèrement. Ainsi naissaient les gammes bien tempérées, dont le cas particulier du tempérament égal où toutes les quintes sont égales et correspondent au rapport

$$2^{7/12} \simeq 1.4983.$$

Dans le tempérament égal, la quinte Ré-La (Ré à 294Hz) produira un battement de

$$(3 \times 1 - 2 \times 2^{7/12}) \times 294\text{Hz} \simeq 1Hz, \tag{3.2}$$

ce qui est bien plus acceptable que celui de 11Hz produit par la quinte Ré-La de Zarlino. Le demi-ton tempéré correspond quant à lui à un rapport de

$$2^{1/12} \simeq 1.0595.$$

Dans cette gamme, tous les demi-tons sont égaux, et il n'y a pas de différence entre les ♯ et les ♭, une même touche du clavier servant à jouer Do♯ et Ré♭.

En fait, de nombreux tempéraments intermédiaires entre celui de Zarlino et le tempérament égal furent en usage à l'époque de Bach. On trouvera dans [7] une description de plusieurs tempéraments historiques, dont le suivant conseillé pour jouer le clavecin bien tempéré de J. S. Bach. Les tonalités y sont d'autant plus consonantes qu'elles sont proches de la tonalité de Do majeur. En partant du Do, on effectue un cycle de 4 quintes Do-Sol-Ré-La-Mi égales, de rapport légèrement inférieur à 3/2, de sorte que la quinte Do-Sol batte par défaut à la même vitesse que la tierce majeure Do-Mi bat par excès : c'est le clavier « bien tempéré». Si l'on note x le rapport de ces quintes, cela donne les battements

$$\text{battement Do-Mi par excès : } b_1 = |5 - 4x^4/4| = x^4 - 5,$$
$$\text{battement Do-Sol par défaut : } b_2 = |3 - 2x| = 3 - 2x,$$

et la condition $b_1 = b_2$ donne

$$x^4 + 2x - 8 = 0$$

dont la solution qui nous intéresse est

$$x \simeq 1.4960.$$

La quinte Si-Fa$\sharp$ est accordée avec le même rapport. Toutes les autres quintes sont accordées justes, c'est-à-dire avec le rapport 3/2. Les octaves sont bien sûr accordées également justes. Ceci est rendu possible du fait que l'on a effectivement

$$x^5 \times (3/2)^7 = 128.0057... \simeq 128.$$

Ce tempérament comporte donc des tierces majeures plus ou moins proches de la tierce harmonique, la plus proche étant celle de la tonalité « principale» Do majeur, et elles vont en s'agrandissant lorsqu'on s'en éloigne. Chaque tonalité aura ainsi sa sonorité propre, au contraire de ce qui se passe avec le tempérament égal.

Ces développements théoriques que nous venons de faire, et qui concernent plutôt les instruments à clavier (dont le son est fixe), partaient de l'hypothèse que les cordes produisent des harmoniques, ce qui n'est pas tout à fait exact comme nous l'avons vu au § 2.5.1. La réalité est donc en fait plus complexe, et dans la pratique, les accordeurs font en sorte que les quintes et les octaves battent légèrement, le moins possible, par défaut pour les quintes et par excès pour les octaves. En poussant même un peu plus loin, l'accordeur S. Cordier [1] a récemment proposé d'accorder les quintes sans battements, ce qui amène des octaves légèrement élargies (cf. aussi § 4.2.2). Sans tenir compte de l'inharmonicité, cela donne un rapport d'octave x devant satisfaire $(3/2)^{12} = x^7$, soit $x \simeq 2.0039$. Le battement produit au niveau de l'octave Do-3 Do-4 est alors de

$$|2 - 2.0039| \times 262 \simeq 1\text{Hz}.$$

Autrement dit, le battement de 1Hz de la quinte (3.2), associé à une octave juste, a été reporté sur l'octave. Pour certaines compositions comme les sonates de Beethoven où les doubles voire triples octaves sont fréquentes, cela peut se révéler désagréable.

En ce qui concerne la pratique instrumentale des musiciens qui ajustent eux-mêmes la hauteur de la note pendant le jeu (violons, vents), le tempérament égal est loin d'être la règle. Si les quintes sont jouées justes, les instrumentistes font en général la différence entre un Do$\sharp$ et un Ré$\flat$, avec une nette tendance à raccourcir les demi-tons des notes sensibles. E. Leipp rapporte dans son ouvrage [10] que le Si, sensible de la tonique Do, est parfois joué à moins d'un quart de ton du Do, sans que cela gêne l'auditeur ! Quant au chant, il est parfois difficile par l'analyse de déterminer quelle est la hauteur exacte d'une note. Sur l'extrait de *la Norma* par Maria Callas présenté sur la figure 1.18 où le diapason est à 441Hz, l'analyse sur 1.8s du La-4 (théoriquement 882Hz) à $t = 7$s a donné la figure 3.4, sur laquelle la hauteur moyenne semble plutôt être autour de 900Hz... Mais, c'est Maria Callas !

3.3.1 Petite histoire du La-3

C'est au luthier anglais John Shore que l'on doit l'invention vers 1711 du diapason à fourche, plus précis que les « flûtes d'accord» utilisées auparavant. Le La-3 du diapason a

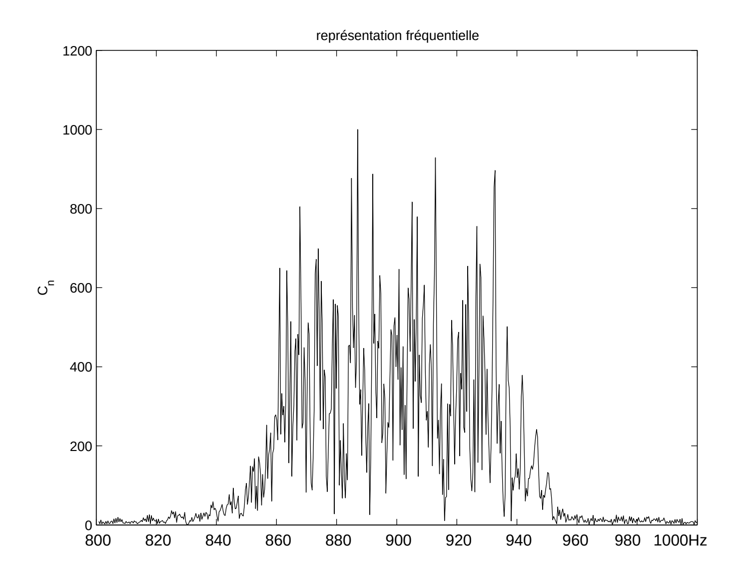


FIG. 3.4 – analyse fréquentielle du La-4 par M.C. (pour ceux qui ont suivi !) où l'on mesure bien la distance qu'il y a entre les modèles simples que nous décrivons et la complexité de la réalité...

beaucoup varié au cours des siècles, mais également d'un endroit à l'autre, ce qui n'allait pas sans poser quelques problèmes aux musiciens qui voyageaient ! Selon certains, le diapason utilisé dans une ville dépendait pas mal de l'état de ses finances : si les fonds venaient à manquer, on rognait sur le budget « orgue» en raccourcissant légèrement les tuyaux, et le diapason grimpait. Les autres instruments devaient alors suivre le mouvement... Le diapason de l'opéra de Paris de 1704 était à 405.3Hz (plus bas que le La♭ actuel), puis il monta graduellement jusqu'à 449Hz en 1858. Au même instant, il était à 434Hz à Londres et 455.5Hz à Bruxelles. Aussi, un décret de 1859 le fixa en France à 435Hz, ce qui fut confirmé par le congrès de Vienne de 1885. Cela ne l'empêcha pas de continuer sa fuite en avant, et l'on retrouve plus tard sa trace dans une conférence internationale réunie à Londres en 1953, qui le porte à 440Hz. Entre-temps, l'Académie Française avait solennellement mais en vain essayé de le faire revenir à son ancienne valeur 435Hz. Aujourd'hui, il se promène en liberté surveillée entre 440 et 444Hz selon les orchestres ! Mais les adeptes de la musique baroque préfèrent jouer sagement au La 415, c'est plus prudent pour les instruments à cordes anciens...

3.3.2 Désignation des notes

Guy d'Arezzo cherchait au début du XIème siècle un système de codification des intervalles. C'est à lui que l'on doit les noms des notes attribués dans les pays latin : Ut (Do), Ré, Mi, Fa, Sol, La (le Si ou Si♭ n'apparaîtra vraiment que plus tard), et il est également à l'origine de la portée. Il proposa comme procédé mnémotechnique un hymne à saint Jean-Baptiste attribué à Paul Diacre (vers 770), dans lequel les six premières notes de la gamme diatonique débutent chaque vers :

UT queant laxis

REsonare fibris
MIra gestorum
FAmuli tuorum
SOLve polluti
LAbii reatum
Sancte Ioannes.

On peut douter aujourd'hui de l'efficacité mnémotechnique du procédé ! Auparavant, on utilisait des lettres pour les notes, usage conservé dans les pays anglo-saxons et allemands :

notation latine	Do	Ré	Mi	Fa	Sol	La	Si♭	Si	Do
notation anglo-saxonne	C	D	E	F	G	A	B♭	B	C
notation allemande	C	D	E	F	G	A	B	H	C

3.4 Autres exemples de gammes

La simplification de la gamme occidentale a sans doute été imposée par la polyphonie. Dans d'autres cultures où la composition musicale s'organise davantage autour de la richesse mélodique et rythmique, les gammes sont bien plus complexes que la nôtre, et comportent une foison d'intervalles de largeur variable, quarts de tons, trois-quarts de tons... dont notre oreille occidentale a bien du mal à distinguer les finesses.
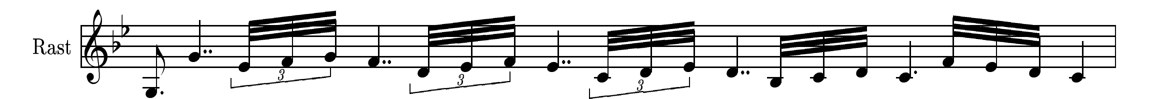


Fig. 3.5 – formules mélodiques sur le maquām *Rast.* Les Si♭ et Mi♭ doivent être joués un quart de ton au-dessus de leur fréquence habituelle.

La notion de gamme est d'ailleurs trop restrictive pour les musiques orientales, et aucun terme ne peut rendre le sens exact du *maquām* de la musique arabe, qui décrit simultanément les intervalles utilisés, le mouvement de la mélodie (ou sa « vie propre» [5]), les points de départ, d'arrêt momentané et de repos final, tout cela organisé généralement à l'intérieur d'un tétracorde (une suite de quatre notes consécutives). Dans le tableau ci-dessous sont indiqués les intervalles utilisés dans les maquām *Rast, Hijāz* et *Saba,* tirés de [5] où vous pourrez trouver des mélodies associées, dont celle reproduite ci-dessus. La notation $\flat^{+}$ signifie ici note baissée d'un quart de ton.

Do	Ré	Mi$\flat^{+}$	Fa	Sol	La	Si♭	Do
Ré	Mi♭	Fa♯	Sol	La	Si$\flat^{+}$	Do	Ré
Ré	Mi$\flat^{+}$	Fa	Sol♭	La	Si♭	Do	Ré (ou Ré♭)

3.5 Sujets d'étude

3.5.1 Fréquences de quelques gammes (***)

Complétez le tableau 3.1. Le Fa le plus grave est Fa-2. Dans tous les cas, on prendra comme fréquence de référence un Do-3 à 261.6Hz, et on supposera que les octaves sont accordées sans battement. Les fréquences seront données avec une décimale de précision. La gamme « bien tempérée» est celle décrite dans le cours pour jouer Bach. Celle avec inharmonicité sera calculée en utilisant la formule du sujet d'étude 2.6.6, et en considérant que B a la même valeur pour toutes les notes : $B = 0.4/1200$, ce qui correspond au cas d'un petit piano droit.

Gamme de :	Fa	Fa♯	Sol	Sol♯	La	Si♭	Si	Do	Do♯	Ré
Pythagore										
Zarlino										
bien tempérée										
tempérament égal										
avec inharmonicité										

Gamme de :	Ré♯	Mi	Fa	Fa♯	Sol	Sol♯	La	Si♭	Si	Do
Pythagore										
Zarlino										
bien tempérée										
tempérament égal										
avec inharmonicité										

TAB. 3.1 – fréquences pour différentes gammes.

3.5.2 Battements des quintes et tierces majeures (*)

Utilisez les résultats du sujet d'étude 3.5.1 pour compléter les tableaux 3.2 et 3.3.

Gamme de :	Fa-Do	Fa♯-Do♯	Sol-Ré	Sol♯-Ré♯	La-Mi	Si♭-Fa
Pythagore						
Zarlino						
bien tempérée						
tempérament égal						
avec inharmonicité						

Gamme de :	Si-Fa♯	Do-Sol	Do♯-Sol♯	Ré-La	Ré♯-La♯	Mi-Si
Pythagore						
Zarlino						
bien tempérée						
tempérament égal						
avec inharmonicité						

TAB. 3.2 – battements des quintes pour différentes gammes.

Gamme de :	Fa-La	Fa♯-La♯	Sol-Si	Sol♯-Do	La-Do♯	Si♭-Ré
Pythagore						
Zarlino						
bien tempérée						
tempérament égal						
avec inharmonicité						

Gamme de :	Si-Ré♯	Do-Mi	Do♯-Fa	Ré-Fa♯	Ré♯-Sol	Mi-Sol♯
Pythagore						
Zarlino						
bien tempérée						
tempérament égal						
avec inharmonicité						

TAB. 3.3 – battements des tierces majeures pour différentes gammes.

3.6 Travaux pratiques sur ordinateur

3.6.1 Construction de quelques gammes

Reprenez le fichier `play.m` vu au chapitre 2 et sauvez-le sous un nouveau nom, par exemple `gammes.m`. Dans ce nouveau fichier, modifiez le vecteur `fr` (tel qu'il est écrit, il contient les fréquences d'une gamme chromatique tempérée légèrement étirée) de telle sorte qu'il contienne les fréquences de la gamme de Zarlino ou de celle de Pythagore, et écoutez la différence entre ces gammes et la gamme tempérée.

3.6.2 Écoute des battements

Nous allons écouter les battements produits entre deux notes. Ouvrez un nouveau fichier, et programmez un accord de deux notes. Pour cela, vous pouvez au choix passer par la fonction `instrument` que vous avez créée lors des TP du chapitre 2 et écrire dans votre fichier (après les initialisations nécessaires) :

```
s = instrument(f1,T,Fe) + instrument(f2,T,Fe) ;
```

ou bien passer par la fonction `synthad` (cf. TP du chapitre 1) en écrivant

```
s = synthad(a,fr1,p,T,Fe) + synthad(a,fr2,p,T,Fe) ;
```

Dans le premier cas, `f1` et `f2` sont les fondamentaux des deux notes de l'accord dont on veut entendre les battements. Dans le second cas, les vecteurs `fr1` et `fr2` contiennent la liste des harmoniques, de la forme `fr1 = (1 :n)*f1` où `n` est le nombre d'harmoniques. On pourra jouer sur les amplitudes contenues dans le vecteur `a`.

Vous testerez les cas suivants :

- `f1` et `f2` très proches ;
- `f1` et `f2` écartées d'une quinte légèrement diminuée. Entendez-vous les battements ? Les entendez-vous encore si vous supprimez les harmoniques de rang 2 et 3 ?
- `f1` et `f2` écartées d'une tierce majeure de Pythagore, ou de Zarlino, ou tempérée. Faites ces tests à différentes hauteurs. Quels sont les harmoniques qui rentrent en jeu dans les battements ? Les entendez-vous encore si vous supprimez ces harmoniques ? (On trouve

là une explication de la difficulté qu'ont les flûtistes à s'accorder en jeu de tierce avec des flûtes très « douces »). Ajoutez-y la quinte pour comparer les accords majeurs de base.

Chapitre 4

Psychoacoustique

La *psychoacoustique* est l'étude de la perception des sons. Le traitement du son effectué par l'oreille et le cerveau est extrêmement complexe, et son étude [15] est rendue délicate par l'intervention de la subjectivité. Nous n'en traiterons ici que quelques aspects : l'intensité et la hauteur qui intéressent directement le musicien, et les effets de *masquage* dont l'étude est très utile pour la mise au point de techniques de compression audio, comme le célèbre format MP3[1] que nous étudierons au chapitre 5.

4.1 Intensité sonore et sonie

L'intensité sonore L_I que nous avons définie au chapitre 1, exprimée généralement en décibels, est une mesure physique de la pression acoustique. Cependant, cette mesure ne coïncide pas avec notre *sensation de l'intensité sonore*, appelée *sonie,* dont l'étude a été développée par Fletcher dans les années 1940. Tout d'abord, nous n'entendons les sons que dans une gamme de fréquences comprises entre 20Hz et 20kHz. Mais même à l'intérieur de cet intervalle, pour un niveau en décibels donné, la sonie n'est pas la même pour toutes les fréquences. L'ouïe présente en particulier un maximum de sensibilité entre 3000 et 4000Hz (cf. fig. 4.1), ce qui permet par exemple au piccolo d'émerger sans efforts d'un *tutti* d'orchestre. Si vous retournez à la figure 1.17, vous remarquerez également que les harmoniques de la soprano sont plus intenses dans cette gamme de fréquences, permettant à sa voix de couvrir facilement l'orchestre. Deux unités de sonie rendent compte de la subjectivité de notre ouïe : le phone et le sone.

4.1.1 Le phone

Par définition, l'intensité en *phones* d'un son pur de 1000Hz, donc de la forme $s(t) = a\cos(2000\pi t + \theta)$, est égale à sa mesure en décibels :

$$x\,\text{dB} = x\,\text{phones à 1000Hz.} \tag{4.1}$$

Ensuite, pour un son pur de fréquence quelconque f, son intensité en phones est par définition *l'intensité en phones du son pur de* 1000*Hz qui produira la même sonie.* Cette nouvelle

[1] raccourci de MPEG Layer 3, qui signifie *moving picture expert group* partie 3 (audio).

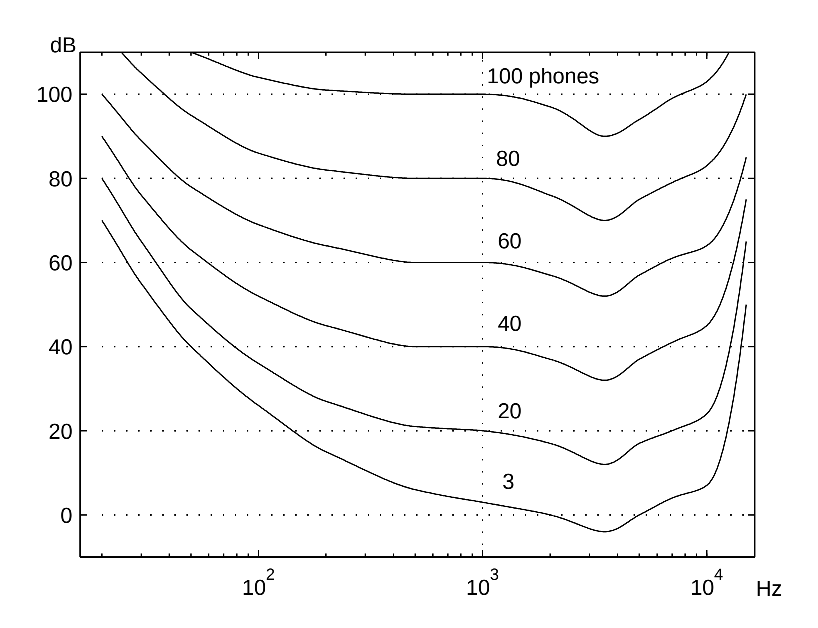


FIG. 4.1 – courbes d'isosonie de Fletcher. Le maximum de sensibilité de l'oreille se situe entre 3000 et 4000Hz.

mesure de l'intensité est notée L_N. Bien entendu, la définition ci-dessus dépend *a priori* de l'auditeur, aussi est-il nécessaire de faire des expérimentations sur un grand nombre de sujets puis d'en faire la moyenne, ce qui a amené l'établissement d'une norme internationale fixant avec précision la relation entre les niveaux L_I (en dB) et L_N (en phones). La figure 4.1 montre quelques courbes d'*isosonie*, c'est-à-dire des courbes le long desquelles la sonie d'un son pur est *constante.* Compte tenu de la définition, une courbe de niveau $L_N = c$ phones passe par le point de coordonnées (1000Hz, c dB). On lit par exemple sur la courbe d'isosonie $L_N = 60$ phones que le son pur de fréquence 100Hz et d'intensité $L_I = 70$dB, ou celui de fréquence 50Hz et d'intensité $L_I = 80$dB, donnent la même sensation d'intensité sonore qu'un son pur de fréquence 1000Hz et d'intensité $L_I = 60$dB. On constate ainsi que les sons très graves (ou très aigus également) nécessitent plus d'énergie pour être perçus au même niveau de sonie. La courbe inférieure $L_N = 3$ phones délimite le seuil d'audition : tout son pur situé en-dessous de cette courbe est inaudible. Vers 2000Hz et 5000Hz, le seuil d'audition correspond à 0dB.

Ces courbes ont été obtenues pour des sons purs. Des études et comparaisons similaires, que nous n'aborderons pas ici, ont été menées pour d'autres types de sons [15] : bruits à bandes variables, sons périodiques, etc.

4.1.2 Le sone

Le phone est relié au décibel par la relation (4.1) et les courbes de Fletcher. Il correspond bien à une mesure de la perception sonore en ce sens que deux sons purs de fréquences différentes qui ont la même mesure en phones seront perçus comme étant au même niveau sonore. Cependant, il ne nous renseigne pas *a priori* sur le niveau en décibels (ou en phones) que devrait avoir un son pour être perçu comme étant deux fois plus fort qu'un autre son.

La loi de Fechner vue au § 1.1.4 énonce bien que la perception de l'intensité suit une loi logarithmique, mais ce point de vue mérite en l'occurrence d'être nuancé [15]. De nombreuses expériences menées sur des sons purs de 1000Hz ont montré qu'en moyenne les sujets trouvent que la sonie est doublée lorsque l'intensité sonore est augmentée de 10dB[2], ce qui correspond à une multiplication de la pression acoustique par $\sqrt{10} \simeq 3.16$ (cf. § 1.1.4). C'est exactement cette relation que traduit la seconde unité de sonie : le *sone*. Un accord international a fixé comme point de référence le son pur de 1000Hz et d'intensité 40dB et lui a attribué une sonie de 1 sone. Le niveau sonore en sones est noté N. Compte tenu des expérimentations mentionnées, le niveau de sonie d'un son pur de 1000Hz est donc

$$N = 2^{(L_N - 40)/10}.$$

Pour trouver ensuite le niveau en sones d'un son pur de fréquence quelconque, il suffit de se reporter sur les courbes d'isosonie de Fletcher. En reprenant l'exemple vu plus haut, le son pur de fréquence 100Hz et d'intensité $L_I = 70$dB ou $L_N = 60$ phones, a donc un niveau de sonie $N = 4$ sones.

En supposant que ce qui précède se généralise aux autres sons, on peut faire les remarques suivantes : dans une salle de concert, le bruit de fond est voisin de 40dB, soit environ 1 sone, et un *tutti* d'orchestre atteint 110dB. Les huit échelons d'intensité *ppp, pp, p, mp, mf, f, ff, fff* correspondent alors plus ou moins à 40, 50, 60, ..., 110dB, soit 1, 2, 4, ..., 128 sones. Nous avons vu au § 1.1.4 que la multiplication par dix du nombre des instruments correspondait à une augmentation de 10dB de l'intensité sonore. Il faut donc multiplier par dix le nombre des instruments (ou par $\sqrt{10}$ leur amplitude) pour multiplier la sonie par deux. Plutôt qu'une loi logarithmique, c'est une loi de type puissance fractionnaire :

$$N \simeq cn^{\alpha}$$

où n est le nombre d'instruments, $\alpha = \log 2 \simeq 0.301$ et c est une constante qui dépend de l'instrument.

4.2 Hauteur des sons

4.2.1 L'oreille

Pour comprendre comment nous arrivons à distinguer entre elles les différentes fréquences, une petite excursion anatomique dans l'oreille s'impose (fig. 4.2 et 4.3). L'élément vibratoire clé est la *cochlée*, fin canal de longueur 32mm environ, rempli de liquide lymphatique, accroché sur l'arête d'une fine lame osseuse et délimité en partie par deux membranes : la membrane basilaire et la membrane de Reissner. Ce canal est tapissé sur sa longueur par quatre rangées comportant chacune environ 3500 capteurs : ce sont les *cellules de Corti*, qui envoient vers le cerveau des impulsions électriques *via* le nerf acoustique. On estime que chaque cellule peut atteindre une cadence maximale de 1000 décharges par seconde, insuffisante pour rendre compte de la discrimination de sons de fréquence supérieure à 500Hz (cf. § 5.1). C'est le fonctionnement conjoint de nombreuses cellules de Corti qui permet une telle discrimination.

[2]Plus précisément, cela vaut lorsque le son le plus faible se situe au-dessus de 40dB. En-dessous de 40dB, l'écart donnant la sensation de doublement croît régulièrement de 3 à 10dB.

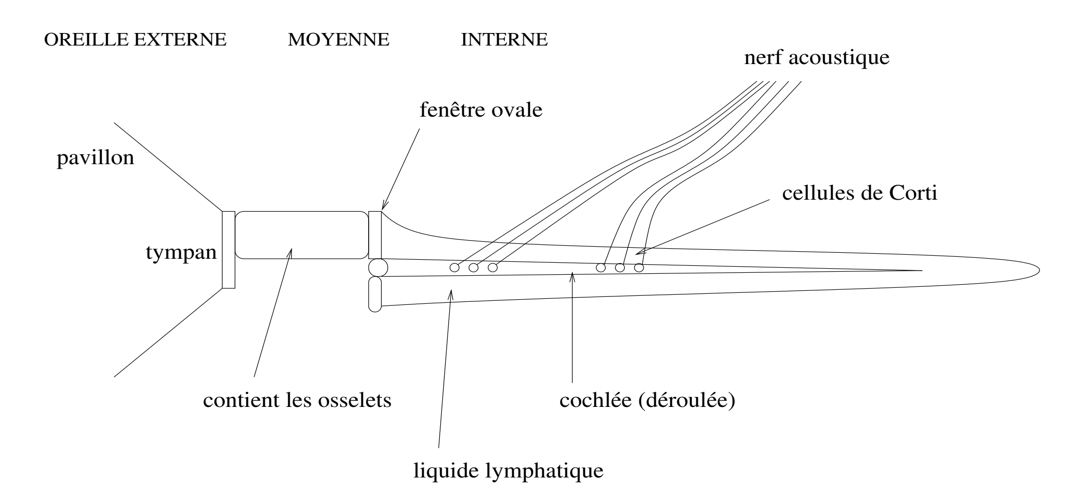


FIG. 4.2 – schéma du fonctionnement de l'oreille. L'organe essentiel, la cochlée, est en milieu aqueux (réminiscence de notre passé de poisson ?). Les osselets servent de levier pour assurer le changement de milieu air-liquide.

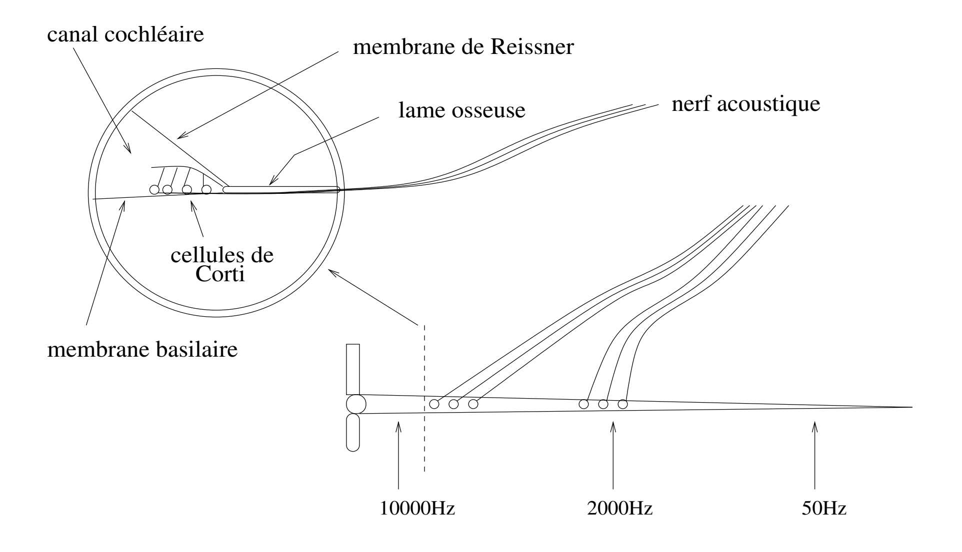


FIG. 4.3 – coupe transversale de la cochlée et localisation des fréquences.

En assignant à chaque impulsion électrique d'une cellule de Corti la valeur 1 (0 correspondant à l'absence d'impulsion), et en supposant que les $4 \times 3500 = 14000$ cellules puissent fournir simultanément 1000 impulsions par seconde, on obtient pour le débit d'une oreille (donc en « mono») la sympathique valeur de 14Mbit/s ! En comparaison, la piste mono d'un CD audio échantillonné à 44kHz sur 16 bits (cf. chapitre 5) a un débit de 0.7Mbit/s, soit vingt fois moins que l'estimation maximale de celui l'oreille. Cela témoigne de la bonne performance de l'oreille, mais aussi du fait qu'augmenter la qualité audio des CD d'un facteur supérieur à 20 ne devrait plus apporter d'amélioration audible.

La section de la cochlée va en diminuant de la fenêtre ovale jusqu'à son extrémité (l'hélicotrême), alors que la membrane basilaire, élastique, va au contraire en s'élargissant, aussi curieux que cela puisse paraître. Le mécanisme complet est difficile à analyser du point de vue mécanique, mais en faisant une observation directe, V. Békésy a constaté vers 1960 que l'amplitude des vibrations de la cochlée dues aux hautes fréquences est maximale dans la partie proche de la fenêtre ovale, alors que pour les basses fréquences, c'est vers l'extrémité que l'amplitude est maximale[15] (cf. fig. 4.3). Ce fonctionnement avait d'ailleurs été prédit en partie au XIXième siècle par le physiologiste et physicien H. Helmholtz.

On retiendra donc qu'à chaque son correspond une « signature sonore» qui se produit sur la cochlée, et qui est envoyée au cerveau par les cellules de Corti. Tous les autres éléments de l'oreille sont là pour assurer le relais entre l'extérieur et la cochlée, puis entre celle-ci et le cerveau :

- L'oreille externe capte les ondes sonores par le pavillon et les achemine au travers du conduit auditif jusqu'au tympan,
- l'oreille moyenne comporte un dispositif chargé de diminuer l'amplitude des vibrations tout en augmentant leur efficacité (effet de levier), assurant la transmission air-liquide des variations de pression : c'est la chaîne des osselets, sollicitée par le tympan, et qui transmet les vibrations à la fenêtre ovale,
- l'oreille interne, en forme de limaçon enroulé sur deux tours et demi environ, est un tube contenant la cochlée qui baigne dans du liquide lymphatique mis en vibration par la fenêtre ovale,
- le nerf acoustique transmet les informations des cellules de Corti vers le cerveau.

4.2.2 Fréquence et hauteur

La sensation de hauteur ou *tonie* est liée bien sûr à la fréquence. En première approximation, en retrouve la loi de Fechner : la tonie varie comme le logarithme de la fréquence. Par exemple, l'écart que nous percevons entre un son de 100Hz et un son de 200Hz est le même qu'entre un son de 200Hz et un son de 400Hz : c'est l'intervalle d'octave, correspondant au doublement de la fréquence. L'intervalle mesuré en octaves entre deux sons purs est donc égal à la différence des logarithmes en base 2 de leur fréquence. Mais ceci ne vaut plus pour les hautes fréquences, et un son pur de 6000Hz paraît être bien loin en-dessous de l'octave d'un son de 3000Hz.

L'échelle des *mels* (échelle de Stevens) a pour but de rendre compte de cette distorsion. Par définition, on attribue 125 mels (ou 131, ou 1000,... selon les sources) à un son pur de 125Hz (*idem*), puis, par des expérimentations sur de nombreux sujets, on étalonne l'échelle des mels de telle sorte qu'un son pur de $2x$ mels donne la sensation d'être exactement une

octave au-dessus d'un son pur de x mels. Par exemple, la suite d'octaves 500, 1000, 2000 mels correspond à peu près aux valeurs 500, 1010, 2050Hz. Toutefois, la pertinence de cette échelle peut être mise en cause pour au moins deux raisons : la très grande variabilité selon les individus, et le fait qu'elle ne vaut que pour des sons purs, à peu près inexistants pour ce qui concerne les instruments acoustiques. Il n'empêche que c'est sans doute là que réside l'explication de la pratique de nombreux accordeurs de piano qui ont tendance à élargir les octaves dans l'aigu, et cela fournit une certaine justification théorique au tempérament à quintes justes de Cordier [1] (cf. § 3.3). Peut-être cela explique-t-il aussi le La-4 de Maria Callas (fig. 3.4) ?

Pour des sons plus complexes, comportant plusieurs harmoniques, la sensation de hauteur n'est pas forcément fournie par la fréquence de l'harmonique présent le plus grave. Pour s'en convaincre, il suffit d'écouter de la musique sur un petit poste radio qui ne délivre aucune fréquence en-dessous de 150Hz par exemple. Ceci n'empêchera pas l'auditeur de reconnaître parfaitement les notes jouées, alors que certains fondamentaux sont absents. En fait, il semblerait que pour identifier une note de fondamental 100Hz par exemple, la présence de quelques harmoniques multiples de 100Hz soit suffisante [10]. Supprimer les harmoniques graves peut toutefois donner la sensation d'un son légèrement plus aigu, alors que supprimer les harmoniques aigus peut donner la sensation d'un son légèrement plus grave. Là réside peut-être l'explication des éternels conflits entre musiciens qui s'accusent de « jouer faux », en particulier lorsque certains jouent dans les coulisses qui, par un effet de filtrage, peuvent en effet supprimer certains harmoniques.

Par ailleurs, il est à noter que la tonie varie également avec l'intensité : les sons graves semblent baisser lorsque leur intensité augmente, alors que les sons aigus semblent monter, la variation apparente de hauteur pouvant atteindre un ton entier lorsque l'intensité sonore passe de 40dB à 100dB, pour des fréquences de 150Hz ou 5000Hz. Pour des sons de 2000Hz, la variation de la tonie est insignifiante.

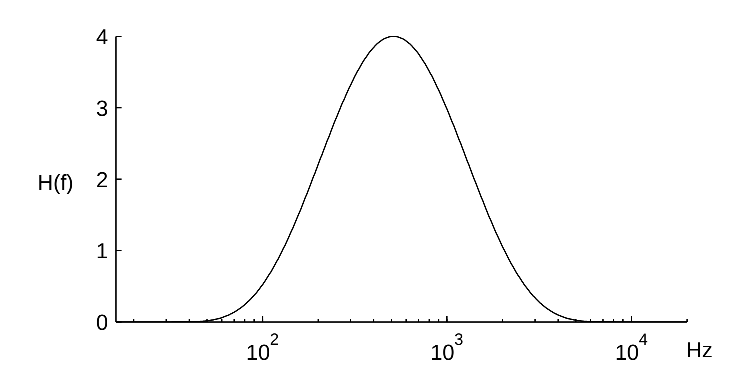


Fig. 4.4 – fonction de troncature des harmoniques.

Terminons ce paragraphe par une belle illusion acoustique sur la hauteur : le son ascendant perpétuel, synthétisé sur ordinateur par Shepard [14], et présentant une certaine analogie avec la reprise dans les orgues (cf. § 2.5.1). L'idée est, sur une gamme ascendante quelconque, par exemple la gamme en tons (une pensée à Claude Debussy), de faire rentrer au fur et à mesure dans chaque note, des harmoniques par le grave en même temps qu'on les fait sortir par

l'aigu.

Pour cela, on se donne une fonction H nulle en dehors de l'intervalle [32Hz, 8192Hz] (les logarithmes des bornes en base 2 sont 5 et 13) et maximale autour d'une fréquence moyenne, par exemple la fonction suivante (cf. fig. 4.4) :

$$H(f) = \begin{cases} (1 + \cos[\pi(2\log_2(f) - 18)/8])^2 & \text{si } 32 \leq f \leq 8192, \\ 0 \text{ sinon.} \end{cases}$$

On considère les fréquences $f_k = 220 \times 2^{k/6}$, $k = 0, 1, 2, \dots$, qui constituent une gamme ascendante en tons. La *gamme ascendante perpétuelle* est alors constituée des sons périodiques suivants (remarquez que les sommes sont finies) :

$$N_k(t) = \sum_{n=-\infty}^{+\infty} H(2^n f_k) \sin(2\pi 2^n f_k t).$$

L'écart entre deux harmoniques consécutifs quelconques de la note N_k est d'une octave. Par ailleurs, les harmoniques de N_{k+1} sont tous situés exactement un ton au-dessus de ceux de N_k, puisque les fréquences correspondantes sont dans un rapport $f_{k+1}/f_k = 2^{1/6}$. Par conséquent, N_{k+1} sera perçu comme étant un ton plus haut que N_k. Or, on a par construction

$$N_6 = N_0\,!$$

Et de même $N_{6+m} = N_m$ pour tout m. On a donc la situation paradoxale suivante : bien que chaque note soit plus haute que la précédente, on revient à la même note six notes plus tard, exactement comme dans le célèbre dessin d'Escher qui représente un escalier en boucle fermée ! Le spectrogramme des huit premières notes est représenté sur la figure 4.5. Sur le même principe, Risset a récemment construit un son ascendant perpétuel « continu», c'est-à-dire en *glissando* [14].

4.2.3 Masquage fréquentiel

Lorsqu'entre deux sons émis ensemble ou presque, l'auditeur n'en perçoit qu'un, on dit que l'autre est *masqué* par celui-ci. Les effets de masquage sont étudiés de manière exhaustive dans [15] : masquage temporel où un son intense masque un son plus faible qui le suit, ou même qui le précède, masquage d'un son pur par un bruit blanc, masquage entre bruits de largeur de bande variable, etc. L'intérêt d'un bon modèle des phénomènes de masquage est de pouvoir en déduire des algorithmes de compression des sons, basés sur le principe simple suivant : *il est inutile de conserver ce que l'auditeur ne percevra pas !* Par conséquent on peut faire passer à la trappe tous les sons masqués, et gagner ainsi sur le volume des données représentant le son. Ceci permet en particulier d'augmenter le débit dans les transmissions par câble ou satellite. Les techniques de compression seront étudiées au chapitre 5.

Nous étudions ici le cas simple mais intéressant du masquage d'un son pur par un autre son pur de fréquence et d'intensité différentes, les deux sons étant émis simultanément. La procédure typique pour mesurer expérimentalement l'effet de masquage est la suivante : l'intensité du son masquant étant maintenue fixe, et partant d'une intensité nulle pour le son masqué, il est demandé au sujet de tourner graduellement un bouton agissant sur son

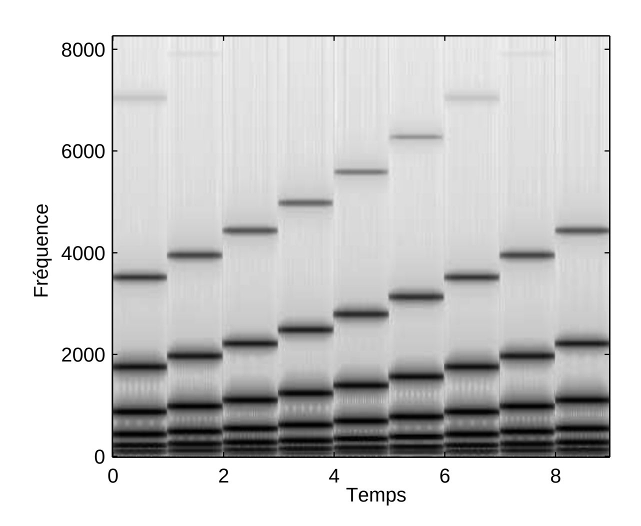


FIG. 4.5 – un éternel recommencement...

intensité jusqu'à ce qu'il devienne audible. En faisant un grand nombre de telles expériences, on obtient une moyenne qui sert de modèle. L'expérience est répétée en faisant varier les fréquences des sons masquants et masqués pour couvrir toute l'étendue des sons audibles.

La figure 4.6 représente le *seuil de masquage* ou *masque* produit par un son masquant de fréquence 1000Hz et d'intensité 80dB. C'est la courbe en trait plein supérieure. Tout son situé en-dessous de cette courbe est masqué, donc inaudible. Pour certains sons, ceux qui sont situés au-dessus de cette courbe mais en-dessous de la courbe en pointillé, le son masqué n'est pas entendu, mais l'auditeur perçoit par contre le *son différentiel* produit par le battement (cf. fig. 2.3) entre les deux sons. Par exemple, si le son masqué a pour fréquence 1200Hz et une intensité de 60dB, le son différentiel entendu a pour fréquence $1200\text{Hz} - 1000\text{Hz} = 200\text{Hz}$. La forme du masque varie à la fois en fonction de la hauteur et de l'intensité du son masquant. La figure 4.7 montre les seuils de masquage pour différentes intensités L_I du son masquant, toujours de fréquence 1000Hz.

Les courbes de seuil de masquage présentent une dissymétrie assez nette : un son grave intense masque assez facilement un son aigu faible, alors qu'un son aigu intense aura plus de mal à masquer un son grave. Une explication possible [21] réside dans la localisation des fréquences (cf. fig. 4.3) : les basses fréquences, pour aller exciter l'extrémité de la cochlée, doivent passer par la zone de réception des hautes fréquences, et donc peuvent avoir une incidence sur la perception de ces dernières. Par contre, les hautes fréquences, localisées au début de la cochlée près de la fenêtre ovale, n'agissent pas sur l'extrémité de la cochlée où sont captées les basses fréquence, et il n'est donc pas surprenant qu'elles n'aient que peu d'incidence sur ces dernières. En se rappelant que les femmes chantent à peu près à l'octave au-dessus des hommes, Zwicker trouve dans cette dissymétrie l'explication du fait que les hommes sont moins nombreux que les femmes dans les chorales. Eh non, ce n'est pas un phénomène socio-culturel !

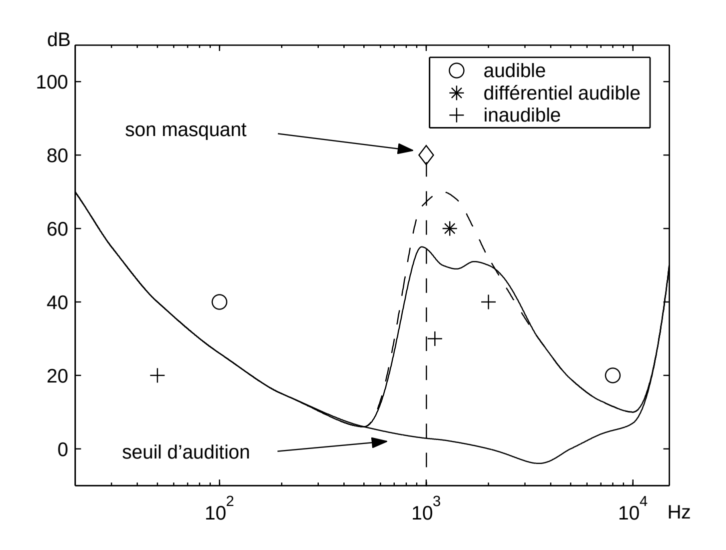


FIG. 4.6 – fréquences masquées par un son pur de 1000Hz à 80dB.

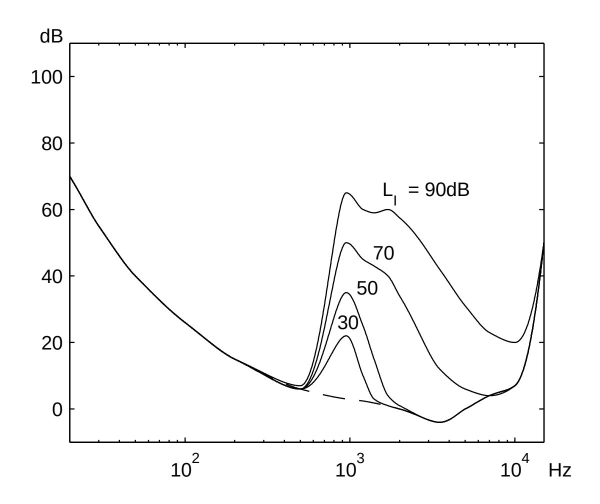


FIG. 4.7 – effet de masque pour un son masquant de 1000Hz, d'intensité variable L_I.

4.3 Sujets d'étude

4.3.1 Niveaux d'isosonie (**)

Les courbes d'isosonie de Fletcher ont été obtenues en faisant la moyenne sur un grand nombre de personnes. Concevez un protocole d'expérimentation et tracez les courbes d'isosonie correspondant à votre propre audition.

4.3.2 Masquage fréquentiel (**)

Concevez un protocole d'expérimentation et tracez les courbes de masquage fréquentiel correspondant à votre propre audition, pour les fréquences masquantes suivantes : 200Hz, 1000Hz et 3000Hz.

4.3.3 Son ascendant perpétuel (**)

En vous inspirant de l'exemple de la gamme ascendante de Shepard, construisez un son qui semble monter en *glissando* indéfiniment, tout en repassant périodiquement par les mêmes valeurs.

4.4 Travaux pratiques sur ordinateur

4.4.1 Masquage fréquentiel

Écrivez un programme qui superpose un son pur de 1000Hz (le son *masquant*) et un son pur de fréquence 1010Hz (le son *masqué*), celui-ci étant successivement d'un niveau sonore relatif de -40dB, -35dB, -30dB, -25dB, -20dB, -15dB par rapport au premier. A partir de quel niveau discernez vous le second son ?

Même questions pour un son masqué de fréquence :
1100Hz et d'un niveau sonore relatif de -40dB, -35dB, -30dB, -25dB, -20dB, -15dB,
1500Hz et d'un niveau sonore relatif de -40dB, -35dB, -30dB, -25dB, -20dB, -15dB,
2500Hz et d'un niveau sonore relatif de -50dB, -45dB, -40dB, -35dB, -30dB, -25dB.

4.4.2 Gamme ascendante perpétuelle

Utilisez les formules du cours pour programmer la gamme en tons perpétuelle. Pour cela, créez une fonction du même format que la fonction `instrument` déjà créée, appelée par exemple `notegp`, et produisant un son s dont la répartition harmonique suit le modèle du cours. Pour créer une suite de $N+1$ notes juxtaposées de ton en ton, vous pouvez écrire une boucle de la forme

```
for k = 0 :N
   f = f0*ton^k ;
   s = [s, notegp(f,T,Fe)] ;
end
```

où `ton` est le rapport qu'il y a entre deux tons consécutifs, et `f0` est la « fréquence» de la première note.

Chapitre 5

Le son numérique

D'une manière générale, on appelle signal *analogique* un signal produit par un dispositif mécanique ou électronique. Dans un tel signal, la variable est le temps qui s'écoule de manière *continue*. Il y a à peine quelques dizaines d'années, toute chaîne de production sonore était entièrement analogique : par exemple, le son produit par les musiciens, le signal électrique délivré par les micros, le signal transmis par ondes hertziennes ou gravé sur un disque de vinyle, le signal reçu et amplifié par votre chaîne Hi-Fi et finalement le son fourni par le haut-parleur, sont tous des signaux analogiques.

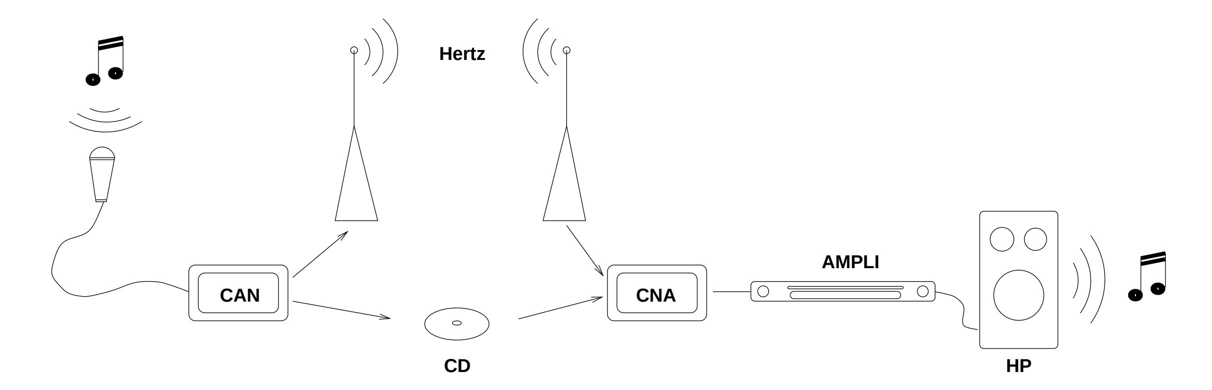


FIG. 5.1 – chaîne audio-numérique

Avec la formidable augmentation de la puissance des ordinateurs, est apparu un nouveau maillon dans cette chaîne : le *son numérique*. Une fois capté par le micro, le son est transformé en une suite de *nombres binaires* (formés de 0 et de 1), qui sont transmis, stockés ou gravés sous cette forme. L'appareil qui assure cette conversion s'appelle un *convertisseur analogique-numérique* (CAN, ADC en anglais pour *analog-to-digital converter*). Il effectue en fait deux opérations distinctes sur un signal analogique $s(t)$:

- l'*échantillonnage*, qui consiste à prélever les valeurs $s_n = s(t_n)$ du signal analogique à des instants régulièrement espacés $t_n = n\tau$, où τ est appelée la *période d'échantillonnage*. La *fréquence d'échantillonnage* $F_e = 1/\tau$ standard pour les CD audio est de 44.1kHz,
- la *quantification*, qui consiste à approcher et remplacer ces nombres réels s_n, qui peuvent

avoir une infinité de décimales impossible à stocker, par des nombres r_n pris dans un ensemble fini comportant $L = 2^b$ valeurs possibles. Ces nombres r_n sont alors codés sur b bits[1] pour être stockés ou transmis. En qualité audio, on utilise généralement un codage sur 16 bits, soit 2 octets.

Le signal obtenu est un signal *numérique*. Il n'a pas d'autre réalité physique que d'être présent sous forme de nombres quelque part dans l'univers informatique, en attendant d'être traité par une carte son. Il est constitué de la suite des nombres r_n, la variable étant maintenant l'entier n : c'est une variable *discrète*. Pour la restitution sonore, le signal numérique est reconverti en signal analogique électrique par un *convertisseur numérique-analogique* (CNA, DAC en anglais), et poursuit son chemin à travers les autres composants classiques de la chaîne.

La question qui vient naturellement à l'esprit est la suivante : quelle est la perte d'information occasionnée par la conversion en signal numérique ? Nous verrons que sous certaines hypothèses, il n'y en a aucune. Ces hypothèses ne sont malheureusement jamais vérifées et ne le seront jamais : il y a là un obstacle théorique fondamental ; mais elles le sont « presque» ! Tout est dans ce presque, imperfection incontournable (mettez ici votre proverbe préféré) mais sans cesse réduite par les poussées de la technologie.

Lorsque l'on cherche à réduire l'erreur introduite par le maillon numérique, le prix à payer réside dans la quantité élevée de données obtenue pour représenter le son : actuellement, une heure d'enregistrement stéréo échantillonné à 44.1kHz sur 2 octets occupera $3600 \times 2 \times 44100 \times 2 = 635$Mo (Mega octets) sur un CD audio. Nous verrons qu'il est possible de réduire la taille des données en utilisant les propriétés psychoacoustiques de l'ouïe, ce qui s'avère très utile, en particulier pour la transmission par internet. Bien entendu, cette compression s'accompagne généralement d'une perte d'information, d'autant plus grande que la compression est poussée. Mais les algorithmes mis en œuvre sont conçus de telle sorte que les données perdues soient le plus possible celles qui justement n'auraient pas été entendues.

Nous terminerons ce chapitre par quelques notions sur le filtrage numérique, et établirons le lien avec le filtrage analogique. Ces notions nous seront utiles au chapitre 6, en particulier pour ce qui concerne les effets sonores.

5.1 Echantillonnage

On considère un son $s(t)$, où la fonction s est continue et bornée sur $\mathbb{R}$. Une période d'échantillonnage $\tau > 0$ ayant été choisie, le *son échantillonné* consiste en la suite des valeurs ou *échantillons*

$$s_n = s(n\tau), \quad n = ..., -1, 0, 1, 2, ...$$

La figure 5.2 représente un son de durée 0.01s, échantillonné à 2000 Hz. À la quantification près (cf. § 5.1.2), les valeurs s_n sont celles qui seront stockées sur un CD audio.

Une description plus élaborée du son échantillonné, abondamment utilisée en théorie du signal, consiste à le représenter sous la forme d'une infinité d'impulsions de Dirac, chacune étant localisée au point t_n et ayant pour *masse* la quantité τs_n, le facteur τ ne servant que de mise à l'échelle (voir équation (5.1)). Nous avons déjà rencontré l'impulsion de Dirac au

[1] Un bit (digit en anglais) vaut 0 ou 1. Un octet (byte en anglais) est composé de 8 bits.

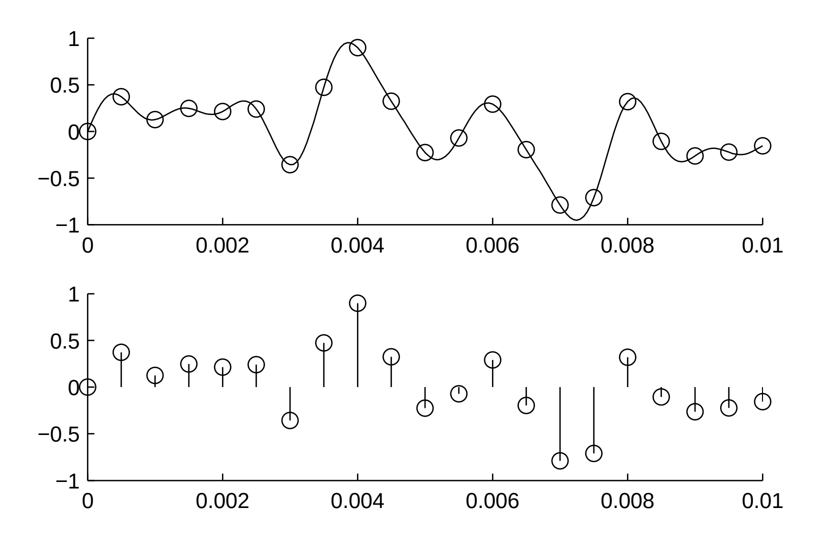


FIG. 5.2 – son initial $s(t)$ (en haut) et son échantillonné $s_e(t)$ (en bas).

point 0 (1.22). De manière générale, l'impulsion de Dirac au point a et de masse $\mu \in \mathbb{C}$, notée $\mu\delta_a$, vérifie la relation

$$\int_{-\infty}^{+\infty} \varphi(t)\mu\delta_a(t)\,dt = \mu\varphi(a)$$

pour toute fonction φ continue et nulle en dehors d'un intervalle borné. Pour mieux se figurer ce qu'est cette impulsion, on peut considérer la suite de fonctions suivantes définies pour $n \geq 1$ (cf. fig. 5.3, où $a = 2$ et $\mu = 1$) :

$$u_n(t) = \begin{cases} \mu n \text{ si } |t-a| < 1/(2n), \\ 0 \text{ sinon.} \end{cases}$$

Ces fonctions vérifient toutes

$$\int_{-\infty}^{+\infty} u_n(t)\,dt = \mu.$$

Ce qui est obtenu à la limite n'est pas une fonction (on l'appelle une *distribution*), mais on a

$$\lim_{n\to\infty} \int_{-\infty}^{+\infty} u_n(t)\varphi(t)\,dt = \mu\varphi(a) = \int_{-\infty}^{+\infty} \varphi(t)\mu\delta_a(t)\,dt.$$

Le son échantillonné est alors défini par

$$s_e(t) = \tau \sum_{n=-\infty}^{+\infty} s_n\,\delta_{n\tau}(t) \tag{5.1}$$

et sa représentation graphique (au facteur τ près) est donnée au bas de la figure 5.2, avec la convention que les cercles ont pour ordonnée la masse de l'impulsion de Dirac associée.

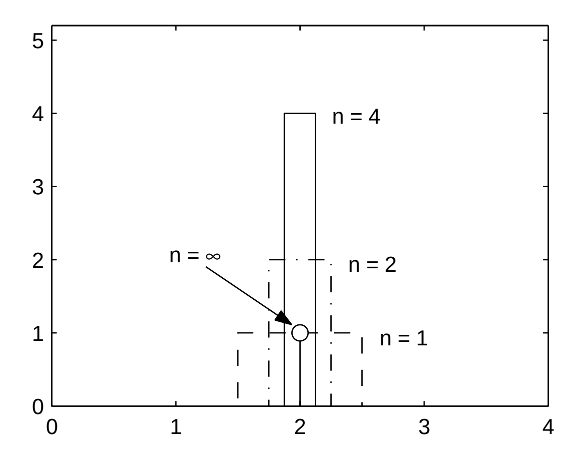


FIG. 5.3 – suite de fonctions $u_n(t)$ dont la limite est l'impulsion de Dirac au point $a = 2$, de masse $\mu = 1$, notée avec le trait vertical et le petit cercle d'abscisse a et d'ordonnée μ.

5.1.1 Critère de Nyquist et théorème de Shannon

La question fondamentale est de déterminer sous quelle condition on pourra reconstruire le son d'origine à partir de la seule connaissance de ses échantillons. À première vue, la tâche semble impossible : on voit bien sur la figure 5.2 que l'on peut modifier les valeurs de $s(t)$ comprises entre deux échantillons sans pour autant changer leur valeur. Par conséquent, deux sons distincts ayant les mêmes échantillons produiront à la sortie du CNA le même signal électrique, puisque le CNA prend en entrée les échantillons et rien d'autre. Au moins l'un des deux signaux reconstruits sera différent du signal de départ !

Pour comprendre quelle limitation il faut imposer au son $s(t)$ pour pouvoir le reconstruire de manière exacte au travers de la chaîne CAN-CNA, commençons par examiner le cas d'un signal sinusoïdal

$$s(t) = \alpha \cos(2\pi f t + \theta).$$

Sachant que l'on a affaire à un signal sinusoïdal, mais dont l'amplitude, la fréquence et la phase sont *a priori* quelconques, quelle est la fréquence minimale d'échantillonnage à imposer pour pouvoir reconstruire $s(t)$ à partir de ses seuls échantillons ? On pourrait proposer de prendre un échantillon à chaque fois que $s(t)$ passe par un maximum, c'est-à-dire un échantillon par période, mais cela ne suffira pas à distinguer un son oscillant d'un son constant (cf. fig. 5.4).

Compte tenu de ce qui précède, il semblerait logique qu'il faille prendre au moins un échantillon à chaque fois que $s(t)$ passe par un maximum ou un minimum, c'est-à-dire *au moins deux échantillons par période*. Sachant que la période vaut $1/f$, cela nous amène à considérer l'hypothèse suivante :

$$\tau < \frac{1}{2f}.$$

La fréquence d'échantillonnage étant $F_e = 1/\tau$, cette condition s'écrit de manière équivalente

$$F_e > 2f, \tag{5.2}$$

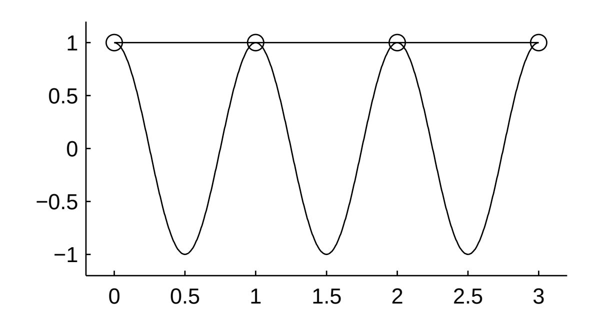


FIG. 5.4 – cosinus et signal constant qui donnent les mêmes échantillons.

que l'on appelle *critère de Nyquist* ou *condition de Shannon*. Pour le raisonnement qui suit, il est plus commode de passer dans le domaine complexe. Rappelons que dans ce cas on doit aussi envisager le cas de fréquences négatives, et le critère ci-dessus devient

$$F_e > 2|f|. \tag{5.3}$$

On se pose donc la question suivante : si deux signaux harmoniques

$$s_1(t) = c_1 \exp(2i\pi f_1 t),$$
$$s_2(t) = c_2 \exp(2i\pi f_2 t),$$

produisent les mêmes échantillons, sont-ils égaux ? De $s_1(0) = s_2(0)$ on déduit d'abord que $c_1 = c_2$, et on pose $c = c_1$. Si $c = 0$, alors $s_1(t) = s_2(t) = 0$ pour tout t, et les deux signaux sont égaux. On suppose maintenant que $c \neq 0$. L'égalité des deux signaux à $t = \tau$ donne

$$c \exp(2i\pi f_1 \tau) = c \exp(2i\pi f_2 \tau),$$

d'où $\exp(2i\pi(f_1 - f_2)\tau) = 1$ et

$$f_1 - f_2 = \frac{k}{\tau} = kF_e, \quad k \in \mathbb{Z}. \tag{5.4}$$

Or, si l'on suppose que les deux fréquences f_1 et f_2 satisfont le critère de Nyquist (5.3), on a

$$|kF_e| = |f_1 - f_2| \leq |f_1| + |f_2| < \frac{F_e}{2} + \frac{F_e}{2} = F_e$$

avec k entier, ce qui impose $k = 0$, et donc $f_1 = f_2$. Par contre, si l'on n'impose pas le critère de Nyquist, k peut être choisi non nul, et dans ce cas $f_1 \neq f_2$. De plus, si l'on observe les échantillons suivants, on constate que

$$s_1(t_n) = c \exp(2i\pi f_1 n\tau) = c(\exp(2i\pi f_1 \tau))^n = c(\exp(2i\pi f_2 \tau))^n = c \exp(2i\pi f_2 n\tau) = s_2(t_n),$$

ce qui fait qu'on a effectivement les mêmes échantillons pour des signaux différents.

Ainsi, le critère de Nyquist est nécessaire et suffisant pour que deux signaux harmoniques qui produisent les mêmes échantillons soient forcément égaux. Ceci est un cas particulier

d'un résultat fondamental en théorie du signal : le théorème de Shannon. Avant de l'énoncer, nous avons besoin de la définition suivante : soit $B > 0$, et $s(t)$ un signal admettant une transformée de Fourier $\widehat{s}(f)$. Rappelons que l'on a (1.14)

$$s(t) = \int_{-\infty}^{+\infty} \widehat{s}(f) \exp(2i\pi ft)\, df,$$

c'est-à-dire que $\widehat{s}(f)$ est la densité de la fréquence f dans le signal $s(t)$. On dit qu'un signal $s(t)$ est *à bande limitée* $[-B, B]$ si

$$\widehat{s}(f) = 0 \text{ pour tout } |f| > B,$$

autrement dit si le signal ne comporte aucune fréquence $|f| > B$. Par exemple, bien que nous n'ayons pas défini la transformée de Fourier de $s(t) = \exp(2i\pi ft)$, ce signal a pour fréquence f et il est donc à bande limitée $[-|f|, |f|]$.

Théorème de Shannon. *Soit $s(t)$ une fonction qui admet une transformée de Fourier $\widehat{s}(f)$ à bande limitée $[-B, B]$. On échantillonne cette fonction à la fréquence F_e. Si F_e vérifie le critère de Nyquist*

$$F_e > 2B,$$

alors $s(t)$ est l'unique fonction à bande limitée $[-B, B]$ qui a pour échantillons les valeurs $(s(n/F_e))_{n\in\mathbb{Z}}$. De plus, si $h(t)$ est le filtre passe-bas idéal (1.26) de fréquence de coupure $F_e/2$, alors on peut reconstruire $s(t)$ à partir du signal échantillonné $s_e(t)$ en lui appliquant ce filtre :

$$s(t) = (h * s_e)(t).$$

Les sceptiques pourront objecter que cela paraît absurde : que se passe-t-il si l'on modifie les valeurs de $s(t)$ entre les échantillons sans changer ceux-ci ? La reconstruction ne pourra pas marcher... Il se passe justement la chose suivante, et c'est ce que dit aussi ce théorème : il est *impossible de modifier* les valeurs $s(t)$ *entre* les échantillons sans modifier ceux-ci, tout en restant à bande limitée $[-B, B]$. La moindre modification fait nécessairement apparaître des fréquences supérieures à B, et le théorème ne s'applique plus. Effectivement, la reconstruction ne marchera pas, mais c'est parce que les hypothèses du théorème ne sont plus satisfaites.

Conséquences

Voyons sur un cas concret les implications de ce résultat. Notre oreille n'entend pas les fréquences supérieures à 20kHz. Aussi, dans une prise de son, il est inutile d'enregistrer les fréquences supérieures à 20kHz. Ceci peut être réalisé (à peu près, cf. paragraphe suivant) soit par un filtrage passe-bas du signal enregistré, soit tout simplement par le fait que le micro n'est pas assez sensible pour capter les ultra-sons. Nous sommes donc dans les conditions du théorème avec $B = 20$kHz, qui nous dit que si l'on échantillonne ce signal à une fréquence supérieure à $2B = 40$kHz, alors le signal d'origine (ou plus exactement celui dont on avait retiré les fréquences supérieures à 20kHz) peut être reconstruit de manière exacte en filtrant le signal échantillonné par un filtre analogique passe-bas adapté. Ce qui explique pourquoi la fréquence généralement utilisée en *qualité audio* est choisie supérieure à 40kHz.

Impossibilité théorique

Tout cela est fort intéressant, mais il y a un *hic* : à part la fonction nulle, *il n'existe pas* de fonction à support compact (nulle en dehors d'un intervalle borné) et dont la transformée de Fourier soit également à support compact (cf. § 1.4.2). Ah bon ? Et alors ? Alors, dans le théorème de Shannon, on a supposé que $\widehat{s}(f)$ était nulle en dehors de $[-B, B]$. Donc $s(t)$ ne peut être nulle en dehors d'un intervalle borné. Tout le problème est là : le théorème ne s'applique qu'à des signaux ayant une « durée de vie» infinie, ce qui n'est jamais le cas en musique : tout son enregistré a un début et une fin ! Et par conséquent, contient nécessairement des fréquences arbitrairement élevées... Cela nous amène à la question suivante :

Que se passe-t-il si le critère de Nyquist n'est pas satisfaite ?

Voyons cela sur un exemple. Pour une qualité audio intermédiaire, prenons comme fréquence d'échantillonnage $F_e = 30$kHz, et échantillonnons le son pur

$$s(t) = 2\cos(2\pi ft) = \exp(2i\pi ft) + \exp(-2i\pi ft)$$

de fréquence $f = 27$kHz, absolument inaudible.

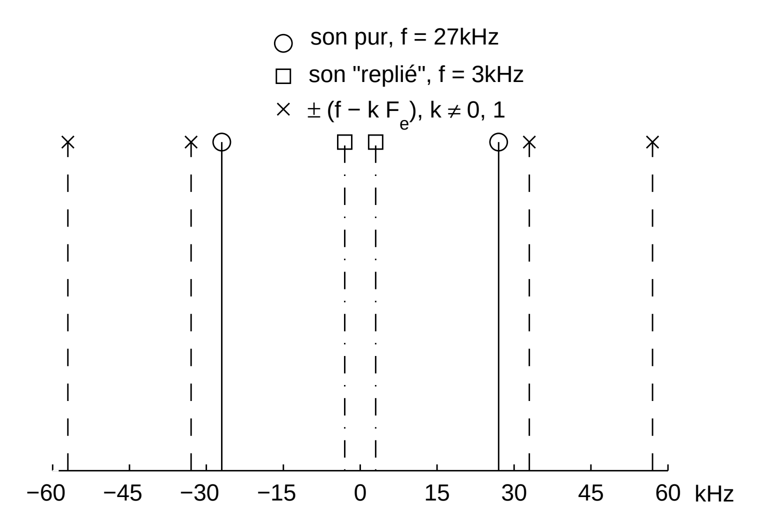


FIG. 5.5 – pour une fréquence donnée f, d'autres fréquences donnent les mêmes échantillons. Attention à celles qui tombent dans la bande $[-F_e/2, F_e/2]$!

D'après l'analyse que nous avons faite au § 5.1.1 et en particulier (5.4), $\exp(2i\pi ft)$ produit les mêmes échantillons que $\exp(2i\pi f_1 t)$ avec $f_1 = f - F_e = -3000$Hz, et $\exp(-2i\pi ft)$ produit les mêmes échantillons que $\exp(2i\pi f_2 t)$ avec $f_2 = -f + F_e = -f_1$. Par simple addition (l'échantillonnage étant linéaire), $s(t) = \exp(2i\pi ft) + \exp(-2i\pi ft)$ a donc les mêmes

échantillons que

$$\begin{aligned} v(t) &= \exp(2i\pi f_1 t) + \exp(2i\pi f_2 t) \\ &= \exp(-2i\pi f_2 t) + \exp(2i\pi f_2 t) \\ &= 2\cos(2\pi f_2 t) \end{aligned}$$

avec $f_2 = 3000$Hz. Cela signifie que $v(n/F_e) = s(n/F_e)$ pour tout $n \in \mathbb{Z}$, autrement dit

$$v_e(t) = s_e(t).$$

D'autres fréquences, de la forme $\pm(f - kF_e)$ donneraient également les mêmes échantillons (cf. fig. 5.5). Prenons alors par exemple $B = 14$kHz. Comme $F_e > 2B$ et le son pur $v(t)$ est à bande limitée $[-B, B]$, on peut appliquer à v le théorème de Shannon, et donc le signal reconstruit par filtrage de s_e, par un filtre passe-bas idéal h de bande passante $[-F_e/2,\ F_e/2]$, sera exactement le son $v(t)$:

$$(h * s_e)(t) = (h * v_e)(t) = v(t).$$

D'un son pur inaudible de 27kHz, on est passé à un son pur de 3000Hz, là où l'ouïe est la plus sensible ! Imaginez l'effet produit par l'intrusion de ce sifflement aigu lors de l'écoute de votre *andante* de Mozart favori...

Ce phénomène s'appelle le *repliement de spectre* (*aliasing* en anglais). Cette terminologie vient du fait que cette fréquence de 3kHz est le point *symétrique* de 27kHz par rapport à la moitié de la fréquence d'échantillonnage $F_e/2 = 15$kHz. C'est comme si l'on avait replié la figure 5.5 le long de l'axe vertical d'abscisse $F_e/2$. Le même repliement a lieu autour de l'axe d'abscisse $-F_e/2$.

Le raisonnement que nous venons de faire vaut pour le cas général. Ce qui se produit dans la réalité, c'est que les sons enregistrés ne sont pas tout à fait à bande limitée [−20kHz, 20kHz] (à cause de l'impossibilité théorique vue plus haut), mais ont toutefois une densité de fréquence faible au-delà de 20kHz.

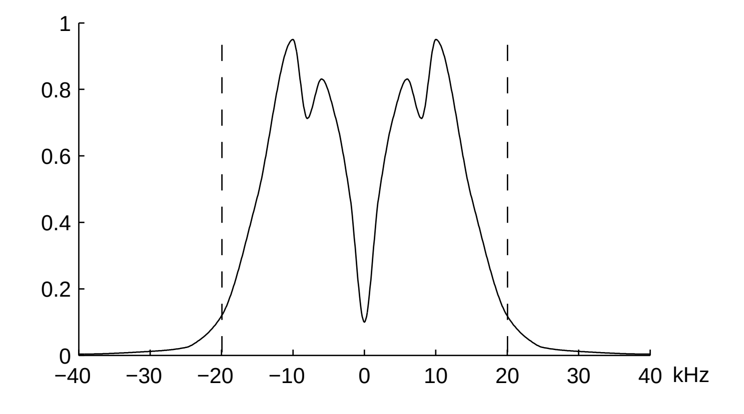


FIG. 5.6 – transformée de Fourier $\widehat{s}(f)$.

La figure 5.6 est une représentation réaliste de ce que peut être la transformée de Fourier d'un tel son : il reste un peu de hautes fréquences à l'extérieur des pointillés délimitant la

bande [−20kHz, 20kHz], mais nous négligeons les fréquences qui sont à l'extérieur de la bande [−40kHz, 40kHz], car cela demanderait des développements théoriques qui dépassent le cadre de ce cours. Supposons que ce son soit échantillonné à la fréquence 40kHz. Un repliement similaire à celui décrit sur l'exemple précédent s'applique également dans ce cas : il s'effectue par rapport à chacun des axes verticaux d'abscisses $-F_e/2 = -20\text{kHz}$ et $F_e/2 = 20\text{kHz}$. Les deux repliements sont représentés en pointillés sur le graphique de gauche[2], figure 5.7. Le résultat du repliement vient se superposer au son initial qu'il perturbe donc légèrement. Ce ce qui sera effectivement entendu après le passage dans le CNA, c'est l'addition du son d'origine avec le repliement, représenté en pointillés sur le graphique de droite de figure 5.7. Dans la partie hautes fréquences de l'intervalle [−20kHz, 20kHz], le son résultant est légèrement enrichi par les contributions des fréquences du son original qui étaient supérieures à 20kHz. Cette perturbation sera d'autant moins perceptible que les valeurs de $|\widehat{s}(f)|$ seront faibles en dehors de l'intervalle [−20kHz, 20kHz].

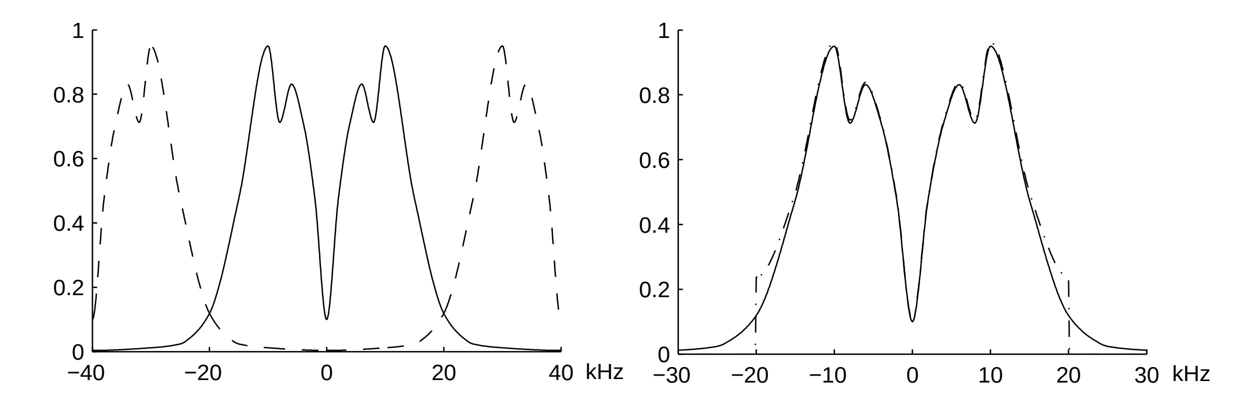


FIG. 5.7 – à gauche : $\widehat{s}(f)$ et les deux repliements. A droite en trait plein : $\widehat{s}(f)$; en pointillé : l'addition des trois spectres, restreinte à l'intervalle [−20kHz, 20kHz]. C'est le son qui sortira du CNA, avec la partie hautes fréquences légèrement altérée.

Pour terminer, voici sur la figure 5.8 le spectrogramme du son ascendant

$$s(t) = \cos(2\pi(1000t + 2000t^2))$$

dont la *fréquence instantanée*, qui augmente progressivement avec le temps, est donnée par la formule

$$f_{\text{inst}}(t) = \frac{d}{dt}(1000t + 2000t^2) = 1000 + 4000t. \tag{5.5}$$

Il s'agit là d'un son synthétique, et les échantillons ont été produits avec MATLAB en calculant les valeurs $s(t_n)$ avec $t_n = n/F_e$ et $F_e = 10000$. Pouvez-vous expliquer ce qui se passe, en particulier à l'instant $t = 1\text{s}$?

[2]En fait, cela vaut uniquement pour la partie paire de $\widehat{s}(t)$, c'est-à-dire sa partie réelle, représentée sur ce graphique. Pour la partie impaire (la partie imaginaire), le repliement s'accompagne d'un changement de signe, comme ce qui se serait produit si nous avions analysé précédemment le cas d'un sinus au lieu de celui d'un cosinus.

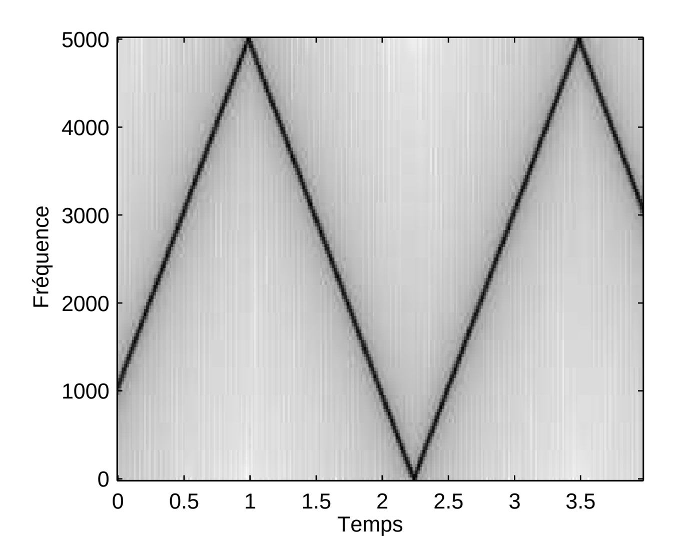


FIG. 5.8 – son ascendant et phénomène de repliement : vous montez ou vous descendez ?

5.1.2 Quantification

Devant l'impossibilité de stocker les nombres réels $s_n = s(n\tau)$, pouvant avoir un nombre illimité de décimales et prendre une infinité de valeurs distinctes, on a recours à la quantification, qui consiste à convertir (coder) chaque nombre réel en un *mot* de longueur finie, pris lui-même dans un ensemble fini. Le *quantum* en question est l'écart minimal qu'il faut entre deux nombres pour être codés différemment. Cette conversion s'accompagne bien entendu d'une perte d'informations irrémédiable, second obstacle à la reconstruction parfaite du son d'origine. Nous décrivons ici le procédé le plus simple de quantification : la *quantification uniforme* [12].

Soient N échantillons s_n, $n = 0, 1, 2, \ldots, N-1$, que l'on souhaite coder en nombres binaires (les mots) sur b bits. Par exemple, dans le cas $b = 3$, ces mots sont 000, 001, 010, 011, 100, 101, 110 et 111, qui sont l'écriture en base 2 des entiers $0, 1, 2, \ldots, 7$. Il y en a en tout $L = 2^b$. Le cas standard $b = 16$ donne 65536 mots. Soit ensuite A une valeur telle que

$$-A \leq s_n \leq A \text{ pour } 0 \leq n \leq N-1.$$

Avec les L mots que nous avons à notre disposition, la quantification *uniforme* consiste à :

- partitionner l'intervalle $[-A, A]$ en L sous-intervalles I_k de *longueur égale* $h = 2A/L$:

  $$I_k = |-A + (k-1)h, -A + kh|, \quad k = 1, 2, \ldots, L.$$

  Nous n'avons pas précisé si l'intervalle est fermé ou ouvert à chaque extrémité, cela restera à préciser lors du codage,
- assigner à chaque échantillon s_n son code binaire b_n, qui est le numéro codé en base 2 de l'intervalle dans lequel il se trouve.

Pour la reconstruction qui suivra, on assigne également à chaque valeur s_n (en fait à chaque b_n) le nombre r_n, milieu de l'intervalle dans lequel il se trouve. La figure 5.9 (qui

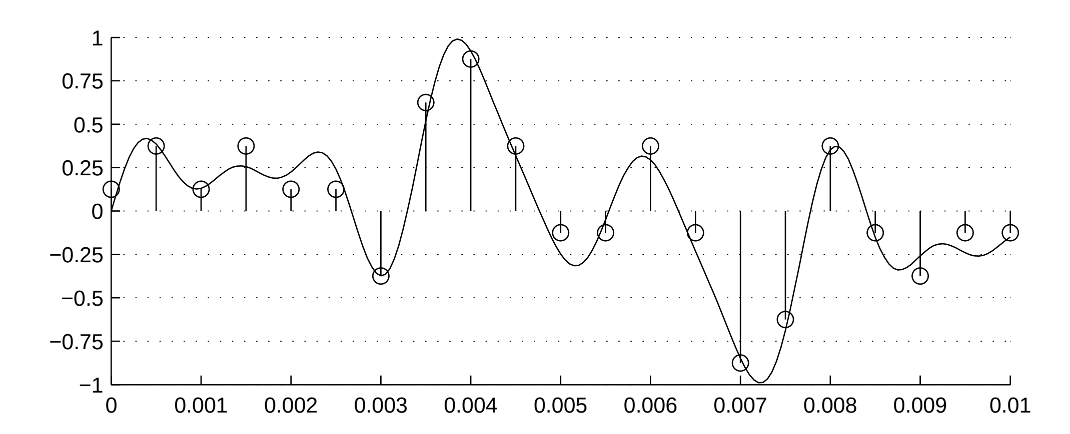


FIG. 5.9 – échantillonnage et quantification sur 3 bits. Il y dans ce cas 8 intervalles. La valeur r_n assignée à un échantillon est le milieu de l'intervalle dans lequel il se trouve, et son code binaire est le numéro de l'intervalle.

reprend l'exemple de la figure 5.2) montre le traitement qui en résulte pour un codage sur 3 bits. On note au passage que la représentation des premiers échantillons n'est pas très satisfaisante. Dans le cas présent, une quantification non uniforme, avec des intervalles plus petits près de 0, aurait probablement donné un meilleur résultat. Le tableau 5.1 donne pour les premiers échantillons, les valeurs de b_n et r_n associées.

s_n	0.000	0.386	0.131	0.255	0.224	0.241	-0.373
r_n	0.125	0.375	0.125	0.375	0.125	0.125	-0.375
b_n	100	101	100	101	100	100	010

TAB. 5.1 – résultat de la quantification : approximation numérique et codage.

Erreur due à la quantification.

Pour chaque échantillon, l'erreur ou *bruit de quantification* $\varepsilon_n = s_n - r_n$ vérifie par construction

$$|\varepsilon_n| \leq \frac{h}{2} = \frac{A}{L} = \frac{A}{2^b}.$$

Le *rapport signal sur bruit* $\mathrm{R_{sb}}$ est la mesure en dB du rapport entre l'intensité sonore (cf. § 1.1.4) du signal I_s et celle du bruit I_b, soit, en prenant pour I_s sa valeur maximale $I_\mathrm{s} = A^2$ et $I_\mathrm{b} = \varepsilon_n^2$:

$$\mathrm{R_{sb}} = 10 \log \frac{I_\mathrm{s}}{I_\mathrm{b}} \geq 10 \log \frac{A^2}{(A/2^b)^2} = 20\, b \log 2 \simeq 6\, b\, \mathrm{dB}.$$

En particulier, *l'ajout d'un bit augmente le rapport signal sur bruit de 6dB*. Pour une *qualité* dite *téléphonique*, on utilise une quantification sur $b = 12$ bits, qui donne un $\mathrm{R_{sb}}$ de 72dB. En qualité audio avec 16 bits de quantification, le $\mathrm{R_{sb}}$ passe à 96dB, ce qui rend le bruit de quantification quasiment imperceptible (cf. chapitre 4). On obtient la même estimation (mais de manière plus rigoureuse) du $\mathrm{R_{sb}}$ en supposant que le signal s_n prend ses valeurs avec

une *probabilité uniforme* sur l'intervalle $[-A, A]$, et que l'erreur ε_n prend ses valeurs avec une probabilité uniforme sur l'intervalle $[-h/2, h/2]$. Dans ce cas, les intensités moyennes respectives de s_n et ε_n (c'est-à-dire leur moyenne quadratique) sont données par

$$\overline{I}_{\mathrm{s}} = \int_{-A}^{A} x^2 \frac{dx}{2A} = \frac{A^2}{3},$$
$$\overline{I}_{\mathrm{b}} = \int_{-h/2}^{h/2} x^2 \frac{dx}{h} = \frac{h^2}{12} = \frac{1}{12}\left(\frac{2A}{2^b}\right)^2,$$

et l'on retrouve la valeur précédente

$$\mathrm{R}_{\mathrm{sb}} = 10 \log \frac{\overline{I}_{\mathrm{s}}}{\overline{I}_{\mathrm{b}}} = 10 \log 2^{2b} \simeq 6.02b\mathrm{dB}. \tag{5.6}$$

Bien entendu, cette valeur ne sera plus la même en cas d'une répartition non uniforme, cas où précisément il peut être approprié d'utiliser une quantification également non uniforme.

5.1.3 Reconstruction du signal sonore

La reconstruction du signal sonore s'appuie sur le théorème de Shannon qui nous dit que si ses hypothèses sont vérifiées, on a

$$s(t) = (h * s_e)(t)$$

où $h(t)$ est le filtre passe-bas idéal (1.26) de fréquence de coupure $F_e/2$. Or ici, nous ne disposons plus des valeurs s_n, mais seulement des valeurs approchées par quantification r_n. Le signal reconstruit sera donc

$$r(t) = (h * r_e)(t) \simeq s(t)$$

où (comparer avec (5.1))

$$r_e(t) = \tau \sum_{n=-\infty}^{+\infty} r_n \, \delta_{n\tau}(t). \tag{5.7}$$

Ici intervient un troisième obstacle à la reconstruction parfaite : on ne sait pas générer un « train» d'impulsions de Dirac $\delta_{n\tau}$. Ici, nous sommes au niveau de la *carte son* (le CNA), chargée de délivrer ces impulsions et d'en filtrer la somme. Tout ce que l'on saura faire, c'est générer de brèves impulsions plus ou moins rectangulaires, comme celles représentées sur la figure 5.3. Cela introduira donc un dernier facteur de distorsion, que l'on pourra partiellement corriger au moyen d'un filtre adéquat. C'est tout l'art des fabricants de cartes son !

5.2 Compression audio

La compression du son numérique est basée sur le principe de ne garder que ce qui est perceptible, et fait donc appel aux propriétés psychoacoustiques étudiées au chapitre 4. Elle vient se placer après la procédure d'échantillonnage et de quantification que nous venons d'étudier. Nous décrivons ici les idées principales mises en œuvre par les algorithmes de compression, dont le plus connu à l'heure actuelle est le codage MPEG Layer-3, en abrégé MP3. Le codage comporte deux étapes distinctes :

- une première étape de *compression psychoacoustique*, accompagnée d'une perte de certaines informations : celles qui sont jugées inutiles au regard des caractéristiques de l'ouïe,
- une seconde étape de *compression entropique*, qui réalise une compression sans pertes du résultat de la transformation précédente.

La description complète de ces deux étapes est assez technique (cf. par exemple [12]), aussi ne ferons-nous qu'en esquisser les grandes lignes. De plus, la première étape, telle qu'elle est implantée dans MP3, fait appel aux *bancs de filtres* utilisés en signal, qui sortent du cadre de ce cours. Nous n'en présenterons donc qu'une version modifiée, mais basée sur le même principe.

5.2.1 Compression psychoacoustique

Dit en deux mots, celle-ci consiste à passer dans le domaine fréquentiel, afin d'effectuer une *quantification des composantes de Fourier* sur un nombre de bits *variable* prenant en compte les propriétés de l'ouïe.

Avant de rentrer dans les détails, on peut faire l'observation suivante, qui éclaire l'utilité du passage au fréquentiel pour les sons musicaux : supposons par exemple que l'on ait un son périodique, échantillonné à la fréquence F_e et quantifié sur 16 bits. D'après (5.6), le $\mathrm{R_{sb}}$ est de 96dB. Une compression temporelle supplémentaire serait par exemple de coder ces échantillons sur 8 bits. On passerait alors à un $\mathrm{R_{sb}}$ de 48dB, et le bruit de quantification deviendrait perceptible : ce n'est plus le son initial que l'on entendrait, mais celui-ci accompagné d'un léger souffle (un bruit blanc). Par contre, si l'on quantifie les coefficients de Fourier sur 8 bits, on introduira certes une légère modification dans les poids relatifs des différents harmoniques, mais on ne fera pas apparaître de fréquences supplémentaires. Il n'y aura donc aucun souffle à la reconstruction du son, mais seulement une modification de timbre, peu ou pas perceptible à ce niveau de quantification. On peut même descendre à 4 bits sans que la perturbation produite soit très gênante, alors que le bruit produit par une quantification temporelle sur 4 bits est carrément insupportable !

Venons-en maintenant aux détails. Notre point de départ servant d'illustration est un son échantillonné à la fréquence $F_e = 22050\text{Hz}$, quantifié sur 16 bits et noté $r_e(t)$. Pour traiter celui-ci, on commence par le découper en petites tranches comportant chacune 512 échantillons. Chaque tranche est obtenue en multipliant le signal par une fenêtre glissante w, comme celle que nous avions utilisée pour le spectrogramme au § 1.4.2, deux tranches consécutives se recouvrant légèrement. La figure 5.10 montre comment est obtenue la deuxième tranche.

Cette tranche u en cours d'analyse comporte donc 512 échantillons u_n codés sur 16 bits, que nous souhaitons compresser. Les propriétés psychoacoustiques que nous allons mettre en œuvre ont rapport aux composantes fréquentielles présentes dans ce signal, aussi la première opération consiste à calculer sa TFD (transformée de Fourier discrète). Les coefficients de cette TFD sont les 512 coefficients complexes $\widetilde{c}_n$ donnés par la formule (1.16), que nous notons ici c_n pour simplifier. Leur intensité

$$L_I(n) = 20 \log |c_n|$$

est représentée en trait plein sur la figure 5.11. Comme dans la figure 1.25, l'abscisse ne

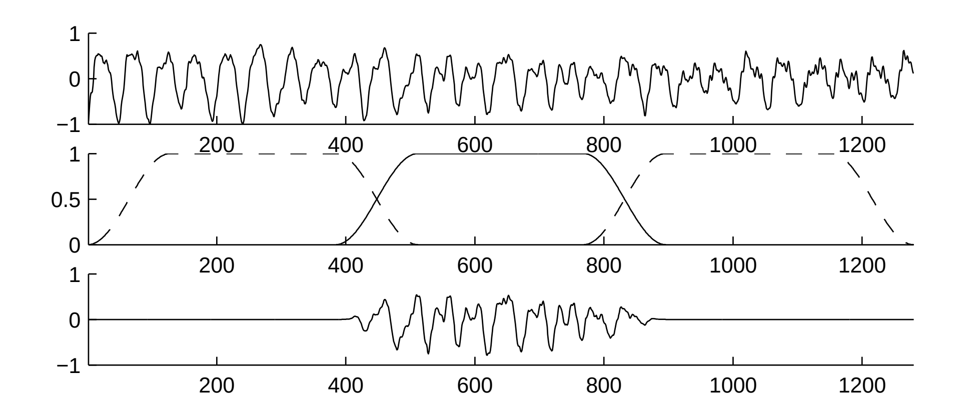


FIG. 5.10 – en haut, le signal échantillonné r (représenté en trait continu pour la lisibilité), extrait de notre air fétiche *la Norma*. Au milieu, trois fenêtres w consécutives, avec en trait plein celle qui est en cours d'utilisation. En bas, la tranche des $u_n = w_n \times r_{n+m}$ à analyser et compresser, avec ici $m = 3 \times 512/4 = 384$, et $n = 0, 1, \ldots, 511$.

représente pas l'indice n, mais la fréquence associée $f_n = nF_e/512\text{Hz}$ où $512/F_e$ est la durée de la tranche analysée.

Nous voyons apparaître en particulier quelques pics fréquentiels, marqués par de petits cercles. Lorsque ces pics ont une intensité située au moins 7dB au-dessus de leurs proches voisins, on les appelle *composantes tonales*. Les autres pics sont appelés composantes *non tonales*. Dans le codage MP3, un traitement différencié est appliqué à ces deux types de composantes, que nous ne décrirons pas ici.

Sur cette même figure apparaissent en pointillé les masques fréquentiels produits par ces pics, ceux que nous avions décrits au § 4.2.3. En l'occurrence, nous avons ici 7 pics qui ont chacun généré un masque $M^{(i)}$, $i = 1, 2, \ldots, 7$. Le masque M représenté sur la figure est *l'enveloppe supérieure* de ces 7 masques et du seuil d'audition S, c'est-à-dire que ses valeurs discrètes M_k sont définies par

$$M_k = \max\left\{M_k^{(1)},\, M_k^{(2)}, \ldots,\, M_k^{(7)},\, S_k\right\}, \quad k = 0, 1, 2, \ldots$$

Ainsi, la partie de la TFD significative pour l'audition est celle située au-dessus du masque M, représentée à gauche de la figure 5.12. Nous notons J les indices des coefficients c_n correspondants. La partie située en-dessous du masque, représentée à droite de la figure 5.12, est *a priori* inaudible car masquée (couverte) par la première. Nous notons K les indices des coefficients c_n relatifs à cette seconde partie.

La compression psychoacoustique consiste alors à *quantifier les coefficients de la TFD, en utilisant moins de bits pour coder les c_k, $k \in K$ que pour coder les c_j, $j \in J$.* Par ailleurs, compte tenu de la formule qui donne les c_n :

$$c_n = \frac{1}{N} \sum_{k=0}^{N-1} u_k \exp(-2i\pi nk/N),$$

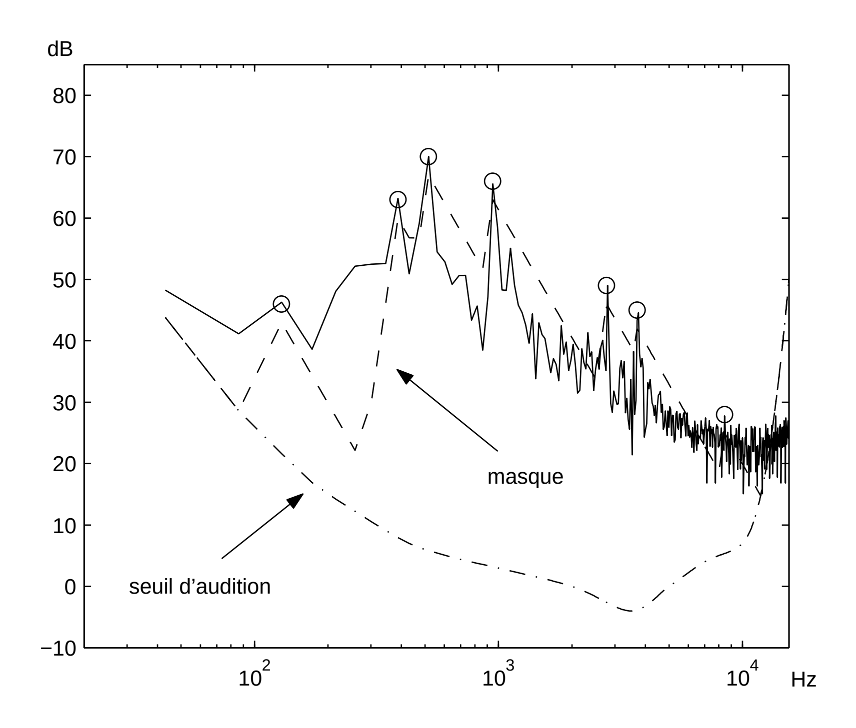


FIG. 5.11 – masque global du spectre (la TFD) des 512 échantillons u_n. Il prend en compte à la fois l'effet masquant des composantes fréquentielles de forte intensité sonore et le seuil d'audition (isosonie 3 phones).

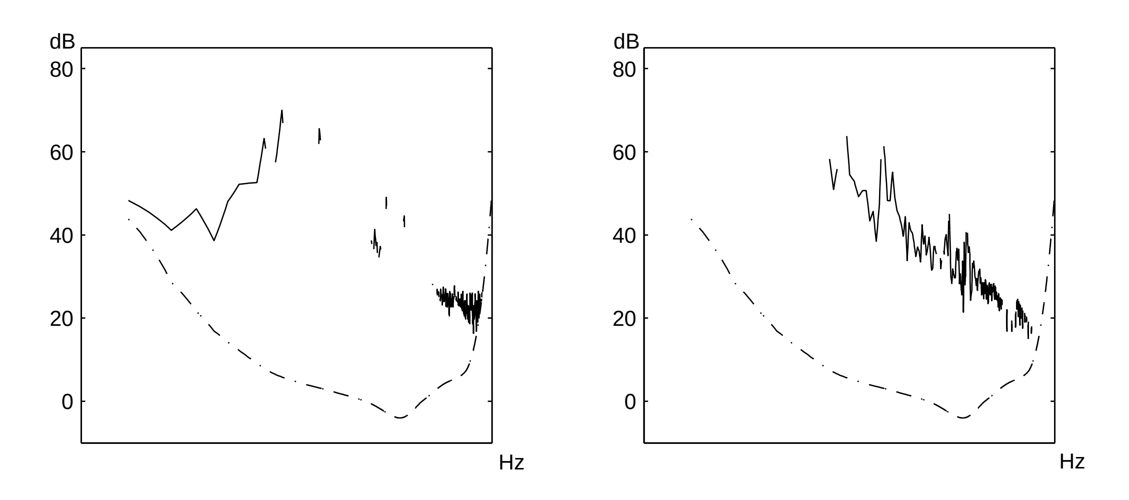


FIG. 5.12 – séparation des parties du spectre situées au-dessus (figure de gauche) et en-dessous (figure de droite) du masque, en vue d'une compression différenciée (nombre de bits variable par coefficient de Fourier).

on constate que l'on a une symétrie conjuguée :

$$c_{N-n} = \overline{c_n}, \tag{5.8}$$

et par conséquent, il suffit de connaître la moitié+1 des coefficients[3] : $c_0, c_1, \ldots, c_{256}$. Ce sont donc ces 257 premiers coefficients complexes (512 coefficients réels car c_0 et c_{256} sont réels) qui seront quantifiés, et stockés après la compression entropique décrite au paragraphe suivant.

Lors de la *décompression* ultérieure (en général au moment de l'écoute), on recalcule l'ensemble des 512 coefficients c_n en utilisant (5.8), on effectue leur TFD inverse (1.17), et on termine par l'addition des tranches successives. Ceci redonnerait exactement les échantillons r_n si l'on n'avait pas codé les c_n sur un nombre restreint de bits. Le son initial est ainsi reconstitué avec une erreur qui en principe n'affecte que des composantes imperceptibles.

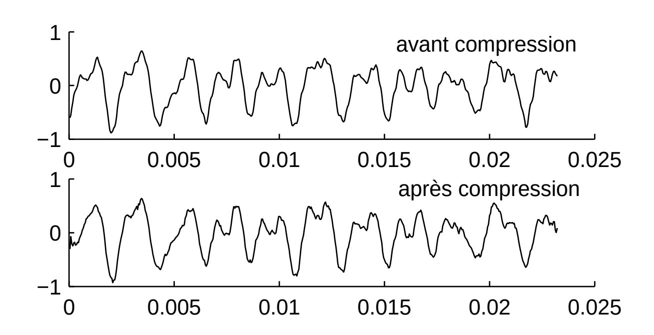


FIG. 5.13 – son initial (en haut), et son après compression-décompression (en bas). Lors de la compression psychoacoustique, la partie de la TFD située au-dessus du masque a été codée sur 8 bits, alors que la partie située en-dessous a été codée sur 2 bits.

Un extrait du son initial et sa reconstitution obtenue par ce procédé sont représentés sur la figure 5.13. L'échantillon sonore initial comportait 512 valeurs codées sur 16 bits, soit en tout 8704 bits. Codée et compressée ensuite par un codage entropique que nous verrons bientôt, la TFD du signal n'occupe plus que 2280 bits, soit un taux de compression de 74%, pour une erreur quadratique relative de 20%. En raffinant (beaucoup) ce procédé, le standard MP3 arrive à des taux de compression de l'ordre de 90% tout en maintenant une excellente qualité sonore.

5.2.2 Compression entropique

A l'issue de la compression psychoacoustique, on dispose d'une suite de nombres ou mots binaires $m_1, m_2, \ldots m_k, \ldots$, de longueur variable. Ce sont les coefficients quantifiés de la TFD. Il arrive fréquemment que l'on ait plusieurs mots consécutifs identiques, aussi commence-t-on par remplacer chaque séquence d'un mot m_k répété p fois par les deux mots $p\ m_k$. On obtient ainsi une deuxième suite de mots binaires, plus courte. Ces mots, qui composent un texte,

[3] Une variante consiste à faire tous les calculs dans le domaine réel en prolongeant par parité les échantillons r_n. Cette variante s'appelle la transformée en cosinus discrète (TCD, DCT en anglais), dans la mesure où une fonction paire se développe en série de cosinus. C'est elle qui est en fait utilisée le plus fréquemment.

font eux-mêmes partie d'un *dictionnaire* comportant un nombre fini de mots d_i. On procède alors à un *changement de dictionnaire*, un peu comme si l'on traduisait le texte dans une nouvelle langue, en s'appuyant sur le principe élémentaire suivant : *coder sur des mots courts les mots les plus fréquents*. Pour cela, on compte le nombre d'occurrences de chaque mot d_i dans le texte, et, en fonction de la fréquence ou *probabilité* de chacun de ces mots, on crée le nouveau dictionnaire en appliquant le principe précédent. Deux exemples de dictionnaires sont donnés dans le tableau 5.2.

	mot1	mot2	mot3	mot4	mot5	mot6	mot7	mot8
dico1	0	10	110	1110	11110	111110	1111110	11111110
dico2	00	01	10	1100	1101	1110	111100	111101

TAB. 5.2 – deux codages possibles des mots (rangés par fréquence décroissante).

Une fois traduits, les mots sont écrits les uns à la suite des autres, sans espace entre eux. Pour que le texte obtenu soit *décodable*, il ne faut donc pas qu'un mot donné puisse être le début d'un autre mot, car cela provoquerait plusieurs interprétations possibles. Par exemple, si les trois mots A, B et C étaient codés 0, 1 et 10, le texte 10 pourrait être interprété aussi bien comme C que comme BA. Cette condition s'appelle la *condition du préfixe*.

Il existe plusieurs stratégies pour construire ce nouveau dictionnaire, mises en œuvre par exemple dans les applications telles que *winzip,* ou dans les commandes `unix` telles que *gzip* ou *compress*. En particulier, il en existe une qui est optimale : l'algorithme de Huffman (cf. par exemple [12]). En reprenant l'exemple précédent, il produit le dictionnaire du tableau 5.3. Bien entendu, lorsque l'on transmet un texte ainsi traduit, il ne faut pas oublier de transmettre le dictionnaire !

	mot1	mot2	mot3	mot4	mot5	mot6	mot7	mot8
probabilité	0.25	0.22	0.19	0.11	0.11	0.05	0.05	0.02
code de Huffman	10	00	110	010	011	1110	11110	11111

TAB. 5.3 – le codage optimal de Huffman.

5.3 Filtrage numérique et transformée en z

Terminons ce chapitre par quelques éléments de *filtrage numérique* qui nous seront utiles au chapitre 6. Pour les signaux analogiques, nous avons vu au § 1.5.2 deux relations fondamentales (1.25) et (1.24) sur le filtrage : si l'on fait passer un signal x dans un filtre de réponse impulsionnelle h (et donc de fonction de transfert $\widehat{h}$), le signal sortant y vérifie

$$y(t) = (h * x)(t),$$
$$\widehat{y}(f) = \widehat{h}(f)\,\widehat{x}(f).$$

Nous allons voir comment se traduisent ces relations du côté de leurs échantillons.

5.3.1 Filtrage numérique

Notons respectivement x_n, h_n et y_n les échantillons des signaux analogiques x, h et $y = h * x$. Nous admettrons le résultat suivant.

Si la condition de Shannon est vérifiée par x ou h : l'un des deux au moins est à bande limitée $[-B, B]$ avec $F_e > 2B$, alors y est à bande limitée $[-B, B]$, et l'on a

$$y_n = \frac{1}{F_e} \sum_{k=-\infty}^{+\infty} h_{n-k} x_k. \tag{5.9}$$

Cette relation définit ce qu'on appelle la *convolution discrète* des signaux $(h_n)_{n\in\mathbb{Z}}$ et $(x_n)_{n\in\mathbb{Z}}$, et l'on conserve la même notation en écrivant[4]

$$y = h * x.$$

On remarquera d'ailleurs la forte analogie avec la convolution continue (1.23) que nous rappelons ici :

$$y(t) = (h * x)(t) = \int_{-\infty}^{+\infty} h(t-u)x(u)\, du.$$

L'intégrale a été remplacée par la somme, et t et u ont été remplacés respectivement par n et k. Ce qu'il est important de noter, c'est que *tout filtrage analogique*, faisant appel selon les cas à des dispositifs acoustiques ou électroniques, *peut être effectué de manière équivalente* (ou à peu près) *dans le domaine numérique*, dès lors que la condition de Shannon est vérifiée (ou à peu près) et que l'on dispose des échantillons de la réponse impulsionnelle du filtre à utiliser.

Lorsque le nombre de coefficients h_n non nuls est *fini*, le filtre est à *réponse impulsionnelle finie,* et on l'appelle filtre RIF (en anglais FIR). Dans le cas contraire, il est à *réponse impulsionnelle infinie,* et on l'appelle filtre RII (en anglais IIR).

5.3.2 Transformée en z

Que se passe-t-il côté Fourier pour les signaux échantillonnés ? Ces signaux ont bien une transformée de Fourier, mais sa définition requiert un outil mathématique qui sort du cadre de ce cours : la théorie des distributions. Nous décrivons ici un autre outil très proche de la transformée de Fourier : la *transformée en z*. Nous verrons que celle-ci permet de représenter de manière très pratique les opérations concernant le filtrage discret.

Définition

Soit un signal discret $x = (x_n)_{n\in\mathbb{Z}}$. On appelle transformée en z du signal x la fonction de la variable complexe z

$$X(z) = \sum_{n=-\infty}^{+\infty} x_n z^{-n}. \tag{5.10}$$

[4]Nous utilisons la même notation pour le signal analogique x et le signal discret représenté par la suite des échantillons x_n. Ceci peut éventuellement prêter à confusion, mais le contexte devrait en principe permettre de déterminer quel est le signal considéré.

Cette somme n'est en général pas définie pour tout z, mais seulement dans une partie du plan complexe, de la forme suivante appelée *couronne* :

$$C(\rho, R) = \{z \in \mathbb{C}\,;\, \rho < |z| < R\},$$

sur laquelle la fonction X est *holomorphe* (qui signifie dérivable par rapport à z).

En pratique, on peut toujours se ramener au cas où $x_n = 0$ pour tout $n < 0$. Un tel signal discret est dit *causal*. Dans ce cas, on a $R = +\infty$ et si $\rho < \infty$, $X(z)$ admet une limite en $+\infty$, le nombre x_0, ce que nous supposons désormais. La question qui se pose alors naturellement est la suivante : étant donné une fonction $u(z)$ qui vérifie ces hypothèses, peut-on lui associer un signal discret x tel que $X(z) = u(z)$ dans la couronne $C(\rho, \infty)$?

Pour y répondre, considérons la fonction $v(z) = u(z^{-1})$, définie pour $|z| < 1/\rho$. D'après la théorie des fonctions holomorphes, elle admet un développement en série de la forme

$$v(z) = \sum_{n=0}^{+\infty} x_n z^n, \quad x_n \in \mathbb{C}.$$

On a alors

$$u(z) = v(z^{-1}) = \sum_{n=0}^{+\infty} x_n z^{-n}. \tag{5.11}$$

Nous avons donc trouvé un signal discret $x = (x_n)_{n\in\mathbb{N}}$ ($\mathbb{N}$ étant l'ensemble des entiers naturels) tel que $X(z) = u(z)$. De plus, par unicité du développement en série de v, ce signal x est *l'unique signal causal* qui a pour transformée en z la fonction $u(z)$.

Effet d'un retard

Si un signal échantillonné x est retardé de m échantillons, ce qui correspond à un retard temporel du signal analogique de $r = m/F_e$ s, on obtient un signal discret y de coefficients

$$y_n = x_{n-m}.$$

La transformée en z de y s'écrit

$$Y(z) = \sum_{n=-\infty}^{+\infty} y_n z^{-n} = \sum_{n=-\infty}^{+\infty} x_{n-m} z^{-n} = \sum_{n=-\infty}^{+\infty} x_n z^{-n-m} = z^{-m} \sum_{n=-\infty}^{+\infty} x_n z^{-n},$$

ce qui donne en définitive

$$Y(z) = z^{-m} X(z). \tag{5.12}$$

Le retard de m échantillons se traduit donc tout simplement par la multiplication de la transformée en z par z^{-m}.

Filtrage et transformée en z

Pour terminer, nous aurons besoin des deux résultats importants suivants, qui nous permettront de manipuler et d'interpréter le filtrage discret.

Soient x et h deux signaux discrets, et $y = h * x$ le produit de convolution défini par (5.9). Soient X, H et Y leur transformées en z respectives, définies sur des couronnes notées C_X, C_H et C_Y. Alors, pour tout $z \in C_X$ et C_H, on a $z \in C_Y$ et

$$Y(z) = H(z)X(z). \tag{5.13}$$

Ici encore, on remarquera l'analogie avec la relation $\widehat{y}(f) = \widehat{h}(f)\widehat{x}(f)$ qui s'applique aux signaux analogiques associés.

D'autre part, compte tenu de l'équivalence avec le filtrage continu décrite au § 5.3.1, et en se plaçant dans l'hypothèse où x est l'échantillonné à la fréquence F_e d'un son analogique vérifiant la condition de Shannon, il est intéressant de pouvoir dire quelle est la fonction de transfert associée au filtrage discret $y = h * x$. Pour ne pas la confondre avec la transformée en z de h, nous noterons cette fonction de transfert $H_t(f)$. Celle-ci est alors reliée à la transformée en z de h par la formule qui suit :

$$H_t(f) = H(\exp(2i\pi f/F_e)). \tag{5.14}$$

Autrement dit, les valeurs de la fonction de transfert sont obtenues en prenant les valeurs de la transformée en z de h *sur le cercle unité* du plan complexe.

L'équivalence avec le filtrage continu se traduit alors de la manière suivante : après reconversion du son numérique $y = h * x$ en son analogique par passage dans le CNA, que nous notons toujours y, celui-ci aura pour transformée de Fourier

$$\widehat{y}(f) = H_t(f)\widehat{x}(f),$$

où x désigne maintenant le son analogique qui avait produit les échantillons x_n. Ceci permet d'interpréter l'action d'un filtrage numérique en considérant les valeurs de $H_t(f)$.

Sous la condition de Shannon, on a $\widehat{x}(f) = 0$ si $|f| \geq F_e/2$. Par conséquent, les seules valeurs de $H_t(f)$ qui nous concernent sont celles pour lesquelles $f \in [-F_e/2,\, F_e/2]$, ou encore $2f/F_e \in [-1,\, 1]$. De plus, en remplaçant x par h et en prenant $z = \exp(2i\pi f/F_e)$ dans (5.10), on a

$$H(\exp(2i\pi f/F_e)) = \sum_{n=-\infty}^{+\infty} h_n \exp(-2ni\pi f/F_e),$$

d'où l'on déduit, en utilisant (5.14), que pour un filtre à coefficients réels h_n on a

$$H_t(-f) = \overline{H_t(f)},$$

et qu'il suffit donc de connaître les valeurs de $H_t(f)$ pour $f \geq 0$. En définitive, les valeurs à considérer sont les $H_t(f)$ pour lesquelles

$$\frac{2f}{F_e} \in [0,\, 1],$$

ce qui est la manière dont sont représentés les filtres discrets dans le logiciel MATLAB. Les quatre modèles de base de filtres sont représentés sur la figure 5.14.

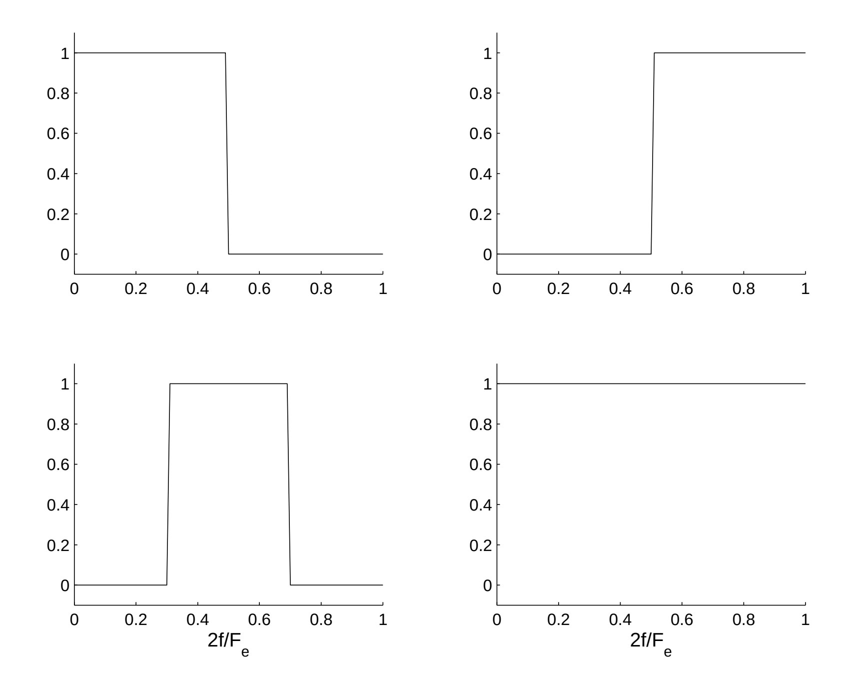


FIG. 5.14 – représentation graphique de $|H_t(f)|$ pour les quatre modèles de filtres les plus courants. La phase de $H_t(f)$ peut varier en fonction de f (non représentée). En haut, à gauche : filtre passe-bas de fréquence de coupure $0.5 \times F_e/2$, et à droite : filtre passe-haut de fréquence de coupure $0.5 \times F_e/2$. En bas, à gauche : filtre passe-bande de fréquences de coupures $0.3 \times F_e/2$ et $0.7 \times F_e/2$, et à droite : filtre passe-tout.

5.4 Sujets d'étude

5.4.1 Critère de Nyquist (*)

Dans le cours, nous avons analysé ce qui se passait si le critère de Nyquist n'était pas vérifié, en prenant comme exemple le son $s(t) = 2\cos(2\pi f t)$, avec $f = 27$kHz et $F_e = 30$kHz. Reprendre cette analyse pour le son $s(t) = \sin(2\pi f t)$.

5.4.2 Repliement d'un son ascendant (*)

Tracez en fonction de f la fréquence effective du son reconstitué après échantillonnage à la fréquence $F_e = 1000$Hz du son $s(t) = \cos(2\pi f t)$. Déduisez-en l'interprétation de la figure 5.8.

5.4.3 Autre cas de reconstruction (***)

Le théorème de Shannon nous dit que si le critère de Nyquist est satisfait, alors la reconstruction d'un son $s(t)$ par filtrage passe-bas idéal du son échantillonné

$$s_e(t) = \tau \sum_{n=-\infty}^{+\infty} s(n\tau)\, \delta_{n\tau}$$

est exacte :

$$s(t) = (h * s_e)\,(t).$$

Nous allons remplacer chaque impulsion de Dirac $\delta_{n\tau}$ par l'approximation vue en cours

$$\delta_{n\tau}(t) \simeq u_\varepsilon(t - n\tau)$$

avec

$$u_\varepsilon(t) = \begin{cases} 1/\varepsilon \text{ si } |t| < \varepsilon/2, \\ 0 \text{ sinon,} \end{cases}$$

où l'on suppose que $0 < \varepsilon < \tau$, τ étant la période d'échantillonnage.
1) En faisant le changement de variable $t' = (t - n\tau)/\varepsilon$, vérifiez que pour toute fonction φ continue sur $\mathbb{R}$, on a

$$\lim_{\varepsilon \to 0} \int_{-\infty}^{+\infty} u_\varepsilon(t - n\tau)\varphi(t)\, dt = \varphi(n\tau) = \int_{-\infty}^{+\infty} \varphi(t)\delta_{n\tau}(t)\, dt.$$

2) Le son échantillonné $s_e(t)$ est remplacé par

$$s_\varepsilon(t) = \tau \sum_{n=-\infty}^{+\infty} s(t) u_\varepsilon(t - n\tau). \tag{5.15}$$

On pose

$$v_\varepsilon(t) = \tau \sum_{n=-\infty}^{+\infty} u_\varepsilon(t - n\tau).$$

Vérifiez que la fonction v_ε est τ-périodique.
3) La fonction v_ε s'écrit donc sous la forme

$$v_\varepsilon(t) = \sum_{k=-\infty}^{+\infty} c_k \exp(2i\pi kt/\tau)$$

où les c_k sont ses coefficients de Fourier. Montrez que

$$\widehat{s}_\varepsilon(f) = \sum_{k=-\infty}^{+\infty} c_k \widehat{s}(f - k/\tau)$$

(on admettra l'interversion $\int \sum = \sum \int$).
4) Le critère de Nyquist étant satisfait, on a $\widehat{s}(f) = 0$ si $|f| > B$ avec $B < 1/(2\tau)$. Représentez

graphiquement sur l'intervalle $[-2/\tau, 2/\tau]$ l'allure de la fonction $\widehat{s}_\varepsilon$ en fonction de celle de $\widehat{s}$. Vérifiez que $c_0 = 1$, et comparez $\widehat{s}(f)$ avec $\widehat{h}(f)\widehat{s}_\varepsilon(f)$, h étant le filtre passe-bas idéal de fréquence de coupure $1/(2\tau)$. Déduisez-en que

$$s(t) = (h * s_\varepsilon)(t).$$

Remarque : dans la réalité, $s(t)$ est remplacé dans (5.15) par $s(n\tau)$, ce qui introduit une légère erreur, d'autant plus faible que ε est petit.

5.4.4 Banc de filtres élémentaire (**)

Étant donné un signal discret $x = (x_n)_{0 \leq n \leq N}$, on considère deux filtres discrets g et h définis respectivement par leurs sorties y et z :

$$y_n = \frac{1}{2}(x_n + x_{n-1}),$$
$$z_n = \frac{1}{2}(x_n - x_{n-1}),$$

avec la convention $x_n = 0$ si $n < 0$ ou $n > N$. Le résultat est ainsi $y = g * x$ et $z = h * x$.
1) Calculez les coefficients g_n et h_n de ces deux filtres.
2) Calculez leur transformée en z, et représentez graphiquement le module de leur fonction de transfert $|G_t(f)|$ et $|H_t(f)|$. Quelle est la nature de chacun de ces filtres : plutôt passe-haut ou passe-bas ?
3) Les sorties y et z sont donc respectivement les composantes basses fréquences et hautes fréquences du signal x. Supposons que l'on veuille utiliser ces données pour compresser le signal. C'est assez mal parti, dans la mesure où l'on a à peu près multiplié par deux le volume des données ! Aussi n'en conserve-t-on qu'un terme sur deux : on pose

$$u_n = y_{2n}, \quad v_n = z_{2n}, \quad 0 \leq 2n \leq N+1,$$

procédé appelé *décimation*. Ces deux signaux u et v sont ensuite compressés par un procédé quelconque, mais sans perte supplémentaire d'informations, que l'on ne discute pas ici. Lors de la phase ultérieure de décompression, l'objectif est de reconstruire le signal de départ x. Pour cela, les signaux décompressés u et v sont re-filtrés par g et h de la manière suivante : on calcule

$$p = 2g * r,$$
$$q = -2h * s,$$

où r et s sont les signaux dits *interpolés*

$$r = (u_0, 0, u_1, 0, u_2, 0, \ldots),$$
$$s = (v_0, 0, v_1, 0, v_2, 0, \ldots).$$

Montrez que $p_{n+1} + q_{n+1} = x_n$ pour $0 \leq n \leq N$. Ainsi, malgré les apparences, aucune information n'a été perdue et x a pu être reconstruit. L'ensemble de ces opérations :

filtrage → décimation → interpolation → filtrage → addition

constitue ce qu'on appelle un *banc de filtres à reconstruction parfaite*.

5.5 Travaux pratiques sur ordinateur

5.5.1 Repliement du spectre

Programmez dans un fichier de commandes le son échantillonné obtenu à partir du son

$$s(t) = \cos(2\pi(1000\,t + 2000\,t^2))$$

en prenant comme fréquence d'échantillonnage $F_e = 11025$Hz et comme durée quelques secondes. Qu'entendrait-on si l'on écoutait le son continu (calculez la fréquence instantanée) ? Écoutez le son numérique obtenu ; que constatez-vous ? Confirmation visuelle : utilisez la fonction `specgram` pour analyser le son.

Chapitre 6

Synthèse et effets sonores

Les premiers à pratiquer la synthèse de sons musicaux ont peut-être été les facteurs d'orgue, qui en juxtaposant plusieurs tuyaux essayaient de reproduire la voix humaine (famille des *régales)* ou les sonorités des instruments de l'orchestre (cf. § 2.5.1). Par exemple, dans *les jeux de gambe*, deux tuyaux jouant ensemble étaient légèrement désaccordés afin de produire un battement, imitant l'ondulation des instrument à cordes : c'est la *voix céleste*, jeu typique du récit romantique. Au cours du XXème siècle, on a assisté à la naissance d'instruments de synthèse électroniques (tout particulièrement des orgues), combinant entre autres circuits oscillants et filtres pour produire des sons musicaux. Aujourd'hui, un avantage remarquable fourni par la technologie du numérique est la possibilité de réaliser sur ordinateur toutes sortes de sons, depuis l'imitation des instruments acoustiques ou électroniques jusqu'à la création de sons entièrement synthétiques, comme le système de synthèse *Modalys* développé par l'IRCAM [18]. Nous verrons quelques procédés simples pour générer de tels sons.

Même s'il est difficile de définir avec précision ce qui les distingue d'un son proprement dit, on peut dire que les effets sonores sont des modifications de sons préexistants. Ils ont sans doute été utilisés de tous temps, et les instruments acoustiques en donnent de nombreux exemples : pédales des clavecins et pianos, sourdines diverses pour la trompette, mais aussi vibrato produit par l'instrumentiste lui-même. Avec l'avènement des instruments électroniques, la musique rock a été une grande utilisatrice d'effets sonores : effet Leslie dans les orgues électroniques, pédales de la guitare (*wah-wah*, sustain, réverbération, saturation), etc. Désormais, la technologie numérique peut reproduire tous ces effets et en créer de nouveaux à l'infini [17]. Par manque de place, nous n'en décrirons que quelques-uns !

6.1 Synthèse de sons musicaux

Les techniques de synthèse numérique de sons musicaux peuvent se classer en deux grandes catégories : il y a celles qui mettent en œuvre une *modélisation physique,* et celles qui utilisent des *modèles de signaux* prédéfinis.

La synthèse par modélisation physique consiste à partir d'un modèle physique simplifié de l'instrument que l'on cherche à produire ou reproduire, et à effectuer ensuite les calculs numériques donnant la réponse du modèle à une excitation donnée. Un exemple élémentaire en est fourni par le sujet d'étude 2.6.5, capable en particulier de rendre compte du phénomène

de rémanence. C'est un domaine que nous ne développerons pas davantage, mais où la recherche est très active [18], et qui commence à apparaître dans les instruments numériques du commerce. La difficulté essentielle de la modélisation physique est de prendre en compte les phénomènes linéaires et non linéaires (au niveau de l'excitateur) de manière à la fois suffisamment simple pour pouvoir effectuer les calculs en temps réel (exigence du musicien), et suffisamment fine pour obtenir un bon rendu sonore.

Dans la synthèse utilisant des modèles de signaux prédéfinis, valable aussi bien dans le domaine électronique que numérique, on peut distinguer au moins quatre types de techniques qui peuvent d'ailleurs se combiner :

- la synthèse *soustractive*, où l'on part d'un son très riche, dont on modèle le spectre par filtrage afin d'obtenir le son désiré (comme le sculpteur qui part d'un bloc plein auquel il retire de la matière),
- la synthèse *additive*, où l'on procède par ajout de sons purs de différentes fréquences, en jouant sur leur pondération et sur leur enveloppe (comme le sculpteur qui procède par ajouts successifs de matière),
- la synthèse par *modulation de fréquence* (FM) inventée par Chowning, inspirée de la technique de transmission par ondes hertziennes,
- la synthèse par utilisation de *sons* préalablement *échantillonnés*.

6.1.1 Synthèse soustractive

La synthèse soustractive nécessite comme point de départ un son riche, mais néanmoins suffisamment structuré pour pouvoir en extraire un son musical, et procède par filtrage de ce son. Elle imite en cela le résonateur d'un instrument de musique qui filtre le signal produit par l'excitateur, ou l'appareil vocal qui filtre le son produit par les cordes vocales, faisant apparaître les formants (fig. 1.28). On prend par exemple comme point de départ un son triangulaire (fig. 1.14 et 2.16) ou rectangulaire (fig. 1.15), et l'on cherche le filtre qui donnera en sortie un son imitant celui des cordes frottées dans le premier cas, des instruments à anche dans le second. Cette technique est toutefois assez gourmande en calculs, et semble avoir eu moins de succès que les autres.

6.1.2 Synthèse additive

Reprenant la démarche des facteurs d'orgue, les concepteurs d'instruments électroniques, puis numériques, utilisèrent la synthèse additive pour créer des sons synthétiques, imitant les instruments acoustiques ou produisant de nouvelles sonorités. Ces sons sont de la forme

$$s(t) = \sum_{n} e_n(t, I) \sin(2\pi f_n t + \theta_n) + b(t, I),$$

que nous avons déjà rencontrée au § 2.5.2 : $e_n(t, I)$ est l'enveloppe de l'harmonique ou partiel de rang n. Ici a été rajouté un bruit $b(t, I)$, qui pourra imiter le souffle d'un instrument à vent, le frottement de l'archet sur la corde, le bruit de percussion au début d'une note de piano, etc. Le paramètre I représente l'intensité à laquelle la note va être jouée : il ne suffit pas en effet de tout multiplier par une même quantité pour obtenir une note plus forte, car l'intensité relative des harmoniques ou partiels peut varier ainsi que leur enveloppe. Le poids

et la forme de $e_n(t, I)$ doivent donc évoluer en fonction de l'intensité I. On s'aperçoit ainsi que dans le cas de l'imitation d'un instrument acoustique, la mise en œuvre s'avère très délicate : il faut avoir au préalable analysé avec précision le son que l'on veut reproduire, à différentes intensités, et déterminer un bon modèle pour faire varier les enveloppes et le bruit en fonction de l'intensité. Cette technique a été utilisée en particulier par Mathews dans son programme Music V pour créer des sons par ordinateur [11, 13]. C'est une méthode qui est également assez gourmande en calculs.

6.1.3 Synthèse FM

La technique de synthèse FM a été mise au point par Chowning en 1973. Rappelons d'abord les trois types de modulation utilisés dans les communications hertziennes : *la modulation d'amplitude (AM), la modulation de phase (PM) et la modulation de fréquence (FM).* Dans les trois cas, on dispose de deux signaux :

- une onde *porteuse,* qui est un signal sinusoïdal de haute fréquence (autour de 1MHz en AM, 100MHz en FM) servant de « véhicule de transport», de la forme $v(t) = \sin(2\pi f_0 t)$ (au déphasage près),
- une onde *modulante* $m(t)$, qui est l'information véhiculée par la porteuse, en général à bande limitée $[-B, B]$ avec $B \ll f_0$.

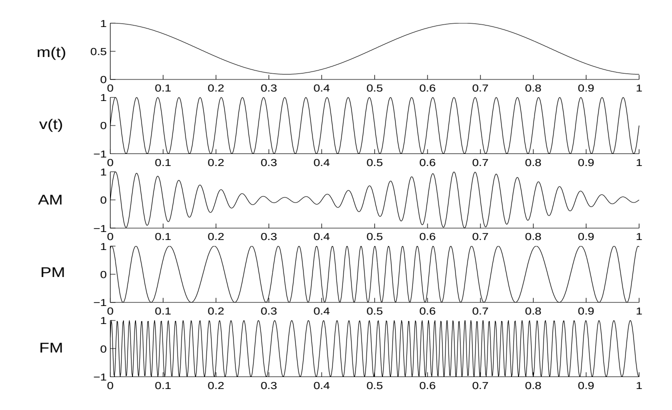


FIG. 6.1 – l'onde modulante, la porteuse et les trois types de modulation. Pour les modulations PM et FM, $m(t)$ et $M(t)$ ont été multipliés respectivement par 20 et 200 afin de rendre visibles les variations dans le signal.

Les trois modulations représentées sur la figure 6.1 correspondent aux signaux suivants :

$$\begin{aligned}
\text{AM} &: s(t) = m(t)\sin(2\pi f_0 t), \quad m(t) \geq 0,\\
\text{PM} &: s(t) = \sin(2\pi f_0 t + m(t)), \quad -\pi \leq m(t) < \pi,\\
\text{FM} &: s(t) = \sin(2\pi(f_0 t + M(t))), \quad M'(t) = m(t).
\end{aligned}$$

En modulations PM et FM, il est important pour la démodulation ultérieure (lors de l'écoute) que $m(t)$ soit à bande limitée $[-B, B]$ avec $0 < B \ll f_0$: la fréquence de la porteuse doit être *nettement supérieure* aux fréquences de la modulante.

L'idée qu'eut alors Chowning en 1973 illustre on ne peut mieux à quel point une idée simple peut révolutionner un domaine. Son idée pour synthétiser des sons fut tout simplement d'utiliser la modulation de fréquence *avec une modulante de fréquence supérieure ou égale à celle de la porteuse !* Autrement dit, les sons qu'il créait ainsi étaient de la forme

$$s(t) = \sin(2\pi f_0 t + \beta \sin(2\pi f_1 t)/f_1)$$

avec $f_1 \geq f_0$. Ici, la fréquence instantanée (cf. (5.5)) est donnée par

$$f_{\text{inst}}(t) = f_0 + \beta \cos(2\pi f_1 t)$$

et $m(t) = \beta \cos(2\pi f_1 t)$. La fréquence de la modulante est f_1, et β est l'*excursion fréquentielle* : la fréquence instantanée varie dans l'intervalle $[f_0 - \beta,\ f_0 + \beta]$. On obtient ainsi toute une gamme de timbres, ne dépendant que des seuls paramètres β et f_1. La figure 6.2 montre la richesse des sons que l'on peut ainsi obtenir, rien qu'en faisant varier f_1.

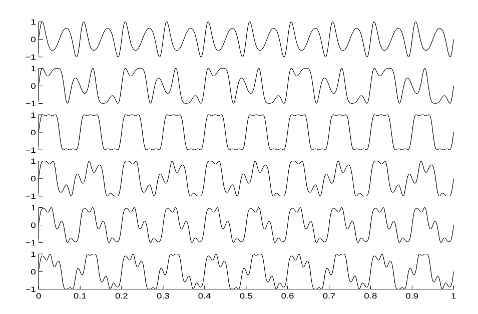


FIG. 6.2 – quelques exemples de sons produits par la technique de Chowning. Ils sont de la forme $\sin(2\pi f_0 t + \beta \sin(2\pi f_1 t)/f_1)$, avec $f_0 = 10$, $\beta = 20$, et de haut en bas : $f_1 = 10, 15, 20, 25, 30, 35$.

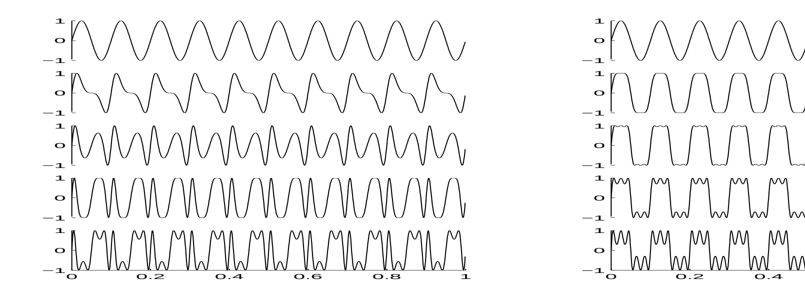


FIG. 6.3 – modification de la richesse harmonique du son en fonction de β, toujours avec $f_0 = 10$. A gauche : $f_1 = 10$, à droite : $f_1 = 20$, et de haut en bas : $\beta = 0, 10, 20, 30, 40$. La ligne 3 se retrouve dans les lignes 1 et 3 de la figure précédente.

Le paramètre β permet quant à lui de modifier le spectre en fonction de l'intensité : pour $\beta = 0$, on a un son pur, qui s'enrichit ensuite en harmoniques aigus au fur et à mesure que l'on augmente β, comme ce qui se passe avec la majorité des instruments à musique (cf. fig. 6.3).

Aussi, lorsque les claviers numériques furent capables de réagir à l'attaque de la touche, on put faire varier le timbre en faisant croître β en même temps que l'attaque.

Cette technique FM offrait de plus une mise en œuvre très simple : il suffisait d'enregistrer les échantillons de la porteuse dans un registre (une mémoire), puis de lire ces échantillons à une vitesse variable. Ainsi naissait la célèbre gamme des synthétiseurs Yamaha DX, dont le très populaire DX7. Par la suite, Chowning utilisa cette même technique pour imiter le chant en reprenant cette fois-ci f_0 multiple entier de f_1, par exemple $f_0 = 17f_1$ ou $f_0 = 25f_1$, imitant les formants de la voix.

6.1.4 Synthèse par utilisation de sons échantillonnés

La synthèse par utilisation de sons préalablement échantillonnés est la technique actuellement la plus répandue pour l'imitation des instruments acoustiques (ou électroniques). Plutôt que de mettre en œuvre l'une des synthèses mentionnées, et dont le résultat est souvent loin de satisfaire les musiciens, on utilise des sons d'instruments préalablement enregistrés et échantillonnés, qu'il suffit alors de restituer au moment voulu. Cela nécessite une capacité de stockage assez importante. Les premiers instruments utilisaient un seul échantillon pour plusieurs notes (une octave par exemple), mais les changements d'échantillon étaient perceptibles : en effet, un Sol d'un instrument quelconque est différent d'un Do dans lequel toutes les fréquences auraient été simplement multipliées par 1.5. Les poids relatifs des harmoniques changent (en général, les poids relatifs des harmoniques élevés vont en diminuant lorsque la fréquence du fondamental augmente). Avec l'accroissement des capacités des mémoires, on en est venu à un puis plusieurs échantillons par note (pour rendre compte également de la variation d'une même note en fonction de son intensité). La tendance actuelle est, à partir d'un modèle physique et d'une base d'échantillons pour chaque note, de réaliser un traitement de ces échantillons pour construire le son en fonction des paramètres d'intensité, de durée, etc.

6.2 Effets temporels : écho et réverbération

Ces effets s'appliquent directement sur le signal temporel. On considère ici un signal discret $x = (x_n)_{n=0}^{N}$ provenant de l'échantillonnage d'un son à la fréquence F_e, de *longueur* finie $N+1$. Deux échantillons consécutifs sont donc séparés par l'intervalle de temps $\tau = 1/F_e$. Par convention, on pose $x_n = 0$ pour $n < 0$ ou $n > N$.

6.2.1 Écho simple

L'*écho simple* est l'effet le plus facile à programmer. Il consiste à rajouter au son initial le même son retardé d'un *délai* r et atténué d'un facteur $0 < g < 1$, qu'on appelle *gain.* On suppose ici que r est un multiple entier de τ : $r = m\tau$. Le signal obtenu y vérifie donc

$$y_n = x_n + gx_{n-m}. \tag{6.1}$$

C'est l'écho qui serait produit par la réflexion du son sur une paroi parfaitement réfléchissante située à une distance $d = cr/2$, avec $g \simeq 1/(2d)$, si l'on suppose que la source et l'auditeur sont placés au même point.

Remarque 6.1 *En utilisant la formule du retard (5.12), la relation (6.1) entre x et y devient par transformée en z*

$$Y(z) = (1 + gz^{-m})X(z).$$

En considérant le polynôme $b(z) = 1 + gz^m$, on a donc $Y(z) = b(z^{-1})X(z)$, ce qui signifie d'après (5.13) que y est le résultat du filtrage de x par le filtre qui a pour transformée en z la fonction

$$H(z) = b(z^{-1}), \tag{6.2}$$

et, d'après (5.11) avec $v = b$, ce filtre a pour échantillons $h_0 = 1$, $h_m = g$, les autres h_k étant nuls. C'est un filtre RIF.

6.2.2 Écho multiple

Examinons maintenant le cas de deux parois qui se font face, et qui engendrent une suite d'échos qui vont en diminuant : c'est l'*écho multiple*. Supposons par exemple que l'auditeur et la source soient situés près de l'une des deux parois. Après une première réflexion sur la paroi opposée, le son reviendra, puis retournera s'y réfléchir une seconde fois, puis une troisième, etc. Appelons y le son ainsi produit. Pour prendre en compte le fait que maintenant le signal provenant de la dernière réflexion n'est plus gx_{n-m} mais gy_{n-m}, la relation (6.1) se transforme alors en la suivante :

$$y_n = x_n + gy_{n-m}. \tag{6.3}$$

C'est une *relation de récurrence* que l'on peut résoudre à la main, ce qui donne

$$y_n = x_n + gx_{n-m} + g^2x_{n-2m} + \ldots + g^kx_{n-km} + \ldots \tag{6.4}$$

Cette relation fait apparaître les réflexions successives, de plus en plus amorties (puisque $0 < g < 1$), mais n'a pas d'intérêt pour le calcul, car le nombre d'opérations à effectuer croît au fur et à mesure que n augmente (environ $3n/m$ opérations pour calculer y_n), alors que dans (6.3), le calcul de y_n nécessite au plus une addition et une multiplication.

Remarque 6.2 *En prenant la transformée en z de (6.3), qui s'écrit également $y_n - gy_{n-m} = x_n$, on trouve cette fois-ci $(1-gz^{-m})Y(z) = X(z)$. En considérant le polynôme $a(z) = 1-gz^m$, on a donc $a(z^{-1})Y(z) = X(z)$, ce qui signifie d'après (5.13) que y est le résultat du filtrage de x par le filtre qui a pour transformée en z la fonction*

$$H(z) = \frac{1}{a(z^{-1})}. \tag{6.5}$$

En appliquant (5.11) avec

$$v(z) = \frac{1}{a(z)} = \frac{1}{1 - gz^m} = \sum_{n=0}^{+\infty} g^n z^{mn},$$

on en déduit que le filtre h a pour échantillons $h_{mn} = g^n$, $n \geq 0$, les autres h_k, k non multiple de m, étant nuls. C'est un filtre RII. Si l'on utilise (5.9) pour calculer $h * x$ avec ces coefficients, on retrouve bien le résultat calculé à la main (6.4).

6.2.3 Réverbération

Dans une salle où il y a plusieurs parois, on entendra de multiples réflexions à des instants variés, qui finiront par se fondre pour produire la *réverbération.* En musique électronique, on obtient une réverbération artificielle en faisant passer le signal (reconverti provisoirement en signal mécanique) dans plusieurs ressorts de longueurs différentes : ce sont les « chambres à écho» (fabriquées en premier par Hammond), que l'on trouve dans la plupart des amplis de guitare. Modéliser une réverbération *numérique* qui soit à la fois peu gourmande en calculs et réaliste est tout un art, et fait l'objet de nombreuses publications (cf. [17] et la bibliographie associée). Nous allons décrire succinctement deux méthodes :

- la première consiste à utiliser la réponse impulsionnelle de la salle, mais elle nécessite un grand nombre d'opérations,
- la seconde superpose des échos simples et multiples en combinaisons avec des filtres passe-tout. Elle est moins gourmande en calculs, mais donne des résultats peut-être moins réalistes.

Utilisation de la réponse impulsionnelle

Nous avons déjà mentionné au § 1.5.2 que le son y perçu par l'auditeur était de la forme

$$y = h * x,$$

où x était le son produit par l'orchestre et h la réponse impulsionnelle de la salle. Ceci concernait des signaux analogiques. Cela reste vrai pour les signaux échantillonnés associés (cf. § 5.3.1), toujours notés x, h et y. Par conséquent, si l'on dispose des échantillons de la réponse impulsionnelle de la salle, que l'on peut obtenir par l'enregistrement d'un son bref approchant l'impulsion de Dirac, on peut obtenir le son discret y en effectuant le produit de convolution des signaux échantillonnés x et h.

Si l'on ne dispose pas de la réponse de la salle, une autre possibilité consiste à l'imiter en utilisant un générateur de nombres aléatoires [17]. Dans la réponse impulsionnelle suivante h, b est un filtre passe-bas, les a_n sont des nombres aléatoires dont la densité suit une loi gaussienne normale centrée, $\tau = 1/F_e$ comme d'habitude, d est une dimension caractéristique de la salle dont on veut simuler la réverbération, et μ est un paramètre supplémentaire de réglage (qui fait d'ailleurs double emploi avec d) :

$$\begin{aligned} h &= b * w, \\ w_n &= a_n \exp(-\mu n \tau / d). \end{aligned}$$

Dans les deux cas, le calcul de la convolution $h * x$ nécessite un grand nombre d'additions et multiplications, ce qui rend cette méthode difficilement utilisable pour les applications en temps réel.

Utilisation d'échos et de filtres passe-tout

Cette technique consiste à combiner les deux types d'échos vus précédemment. La figure 6.4 montre un exemple de structure pour simuler une réverbération. Dans ce qui suit, pour ne

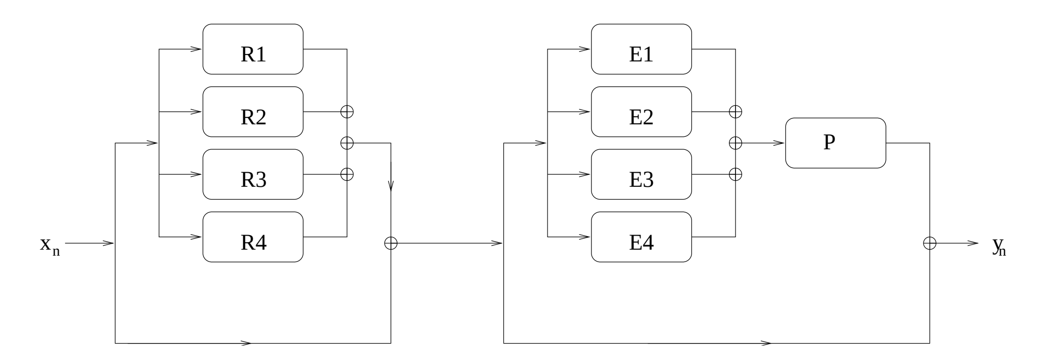


FIG. 6.4 – réverbérateur de Moorer, constitué de lignes à retard R, d'échos multiples E et d'un filtre passe-tout P. Les $\oplus$ représentent des additions.

pas alourdir la notation, nous ne mettons pas d'indices sur les paramètres g, $\mu, \ldots$, associés aux différents éléments R et E, étant entendu qu'ils *peuvent varier* d'un élément à l'autre.

Les éléments notés R sont de simples retards, qui ont pour but de simuler les premières réflexions. Ils délivrent chacun un signal de la forme

$$u_n = gx_{n-m}$$

où $m = rF_e$, r étant le délai qui peut varier par exemple entre 20 et 100ms pour l'ensemble des éléments R. Le coefficient g est l'intensité de l'écho simple, que l'on peut prendre de la forme

$$g = \mu/(cr),$$

où $0 < \mu \leq 1$ représente l'absorption de la salle, et cr la distance parcourue par le son pour revenir à son point de départ après réflexion.

Les éléments notés E sont des échos multiples combinés avec des filtres passe-bas. Un tel élément E, recevant un signal u, délivre un signal

$$w = \mu h * v$$

où μ est toujours un paramètre d'absorption, h est un filtre passe-bas (sinon on obtient un son « métallique») et v vérifie une variante de (6.3) :

$$v_n = u_{n-m} + gv_{n-m}.$$

Chaque gain g associé à un élément E est pris de la forme

$$g = 10^{-3r/T_R}$$

où r est le délai associé, pris entre 50 et 80ms, et T_R est la *durée de réverbération*, qui est par convention le temps que met un son bref réverbéré pour diminuer en intensité de 60dB.

Enfin, le dernier élément noté P est un filtre passe-tout (cf. § 1.5.3). Recevant un signal w, il délivre un signal y qui vérifie l'équation de récurrence

$$y_n = gy_{n-m} + w_{n-m} - gw_n. \tag{6.6}$$

Les paramètres de ce filtre proposés par Moorer [17] sont $g = 0.7$ et $m = 6F_e10^{-3}$ (toujours supposé entier). Ce filtre « brouille les cartes» en modifiant de manière différente les phases de chacune des fréquences, reproduisant l'action d'une salle sur les sons.

Remarque 6.3 *En prenant la transformée en z de (6.6), on obtient ici*

$$Y(z) = \frac{-g + z^{-m}}{1 - gz^{-m}} W(z),$$

et la transformée en z du filtre associé est

$$H(z) = \frac{-g + z^{-m}}{1 - gz^{-m}}.$$

La fonction de transfert de ce filtre est d'après (5.14)

$$H_t(f) = \frac{-g + \exp(-2mi\pi f/F_e)}{1 - g\exp(-2mi\pi f/F_e)}$$

qui, g étant réel, vérifie

$$|H_t(f)| = 1.$$

Ce filtre est donc bien un filtre passe-tout : il ne modifie pas l'intensité des sons purs, il ne fait que changer leur phase.

6.3 Effet par modification du timbre

Nous décrivons ici l'effet « wah-wah» et son implantation numérique. Cet effet fait penser aux formants des voyelles « o» et « a», d'où bien évidemment son nom. Il consiste à rajouter au son initial le son obtenu par filtrage de celui-ci par un filtre passe-bande de bande passante variable : grave pour le son ressemblant au « o», plus aigu pour le son ressemblant au « a» (cf. fig. 1.28).

6.3.1 Un exemple de filtre passe-bande

On peut obtenir un filtre passe-bande numérique demandant peu de calculs par la formule de récurrence

$$y_n = (1 + c)(x_n - x_{n-2})/2 - d(1 - c)y_{n-1} + cy_{n-2}, \tag{6.7}$$

dont la transformée en z est donnée par

$$H(z) = \frac{(1 + c)(1 - z^{-2})/2}{1 + d(1 - c)z^{-1} - cz^{-2}}. \tag{6.8}$$

Rappelons (5.14) que la fonction de transfert du filtre est alors $H_t(f) = H(\exp(2i\pi f/F_e))$. En choisissant les paramètres de ce filtre de la manière suivante :

$$c = \frac{\text{tg}(\pi f_b/F_e) - 1}{\text{tg}(\pi f_b/F_e) + 1},$$

$$d = -\cos(2\pi f_m/F_e),$$

on obtient un filtre passe-bande de largeur de bande f_b centrée sur f_m, c'est-à-dire de fréquences de coupure $[f_m - f_b/2,\ f_m + f_b/2]$. Pour un filtre non idéal, une fréquence de coupure f_c correspond par convention à une *intensité divisée par deux* (soit -3dB) par rapport au maximum dans la bande passante (normalisé à 1), et donc à $|H_t(f_c)| = 1/\sqrt{2}$ (l'intensité étant proportionnelle au carré de l'amplitude, comme nous l'avons vu au chapitre 1). À l'intérieur de la bande passante, on a $|H_t(f)| \geq 1/\sqrt{2}$, alors qu'à l'extérieur $|H_t(f)| \leq 1/\sqrt{2}$. La figure 6.5 représente trois exemples de filtres passe-bande construits sur la relation (6.7).

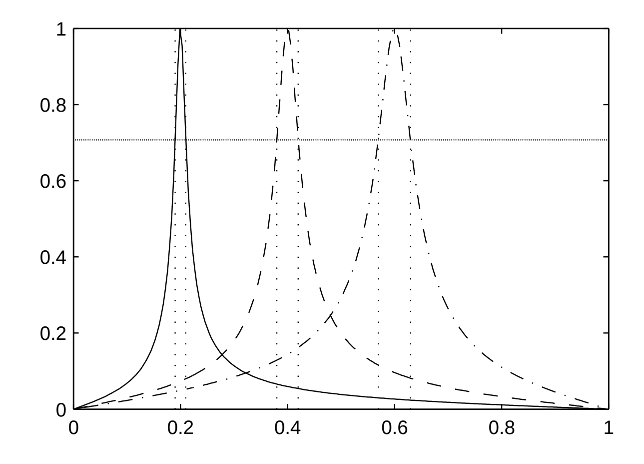


FIG. 6.5 – trois filtres passe-bande, correspondant à $2f_m/F_e = 0.2$, 0.4, 0.6, et $f_b = f_m/10$. Les fréquences de coupures du premier filtre sont $0.19F_e/2$ et $0.21F_e/2$.

6.3.2 Effet « wah-wah»

L'effet *wah-wah* consiste à appliquer un tel filtre passe-bande au son *en faisant varier au cours du temps la fréquence médiane* f_m. La largeur de bande peut être maintenue fixe ou non selon les goûts. Puisque f_m et f_b varient à chaque instant, il en va de même des coefficients c et d de la relation de récurrence (6.7), qui devient

$$y_n = (1 + c_n)(x_n - x_{n-2})/2 - d_n(1 - c_n)y_{n-1} + c_n y_{n-2},$$

avec

$$c_n = \frac{\text{tg}(\pi f_b(n)/F_e) - 1}{\text{tg}(\pi f_b(n)/F_e) + 1},$$
$$d_n = -\cos(2\pi f_m(n)/F_e).$$

Dans les exemples présentés sur les figures 6.6 et 6.7, on a choisi $f_m(n)$ et $f_b(n)$ de la manière suivante :

$$f_m(n) = 2000 + 1000 \sin(2\pi n/F_e),$$
$$f_b(n) = f_m(n)/10.$$

On notera que l'effet *wah-wah* ne modifie pas la hauteur du son (c'est bien visible sur la figure 6.7). Il ne fait que changer les poids relatifs des différentes composantes fréquentielles, à la manière des formants de la voix humaines. Une variante de l'effet *wah-wah* consiste à mettre en parallèle plusieurs filtres passe-bande (ou éventuellement *coupe-bande*) variables, agissant chacun sur une partie du spectre.

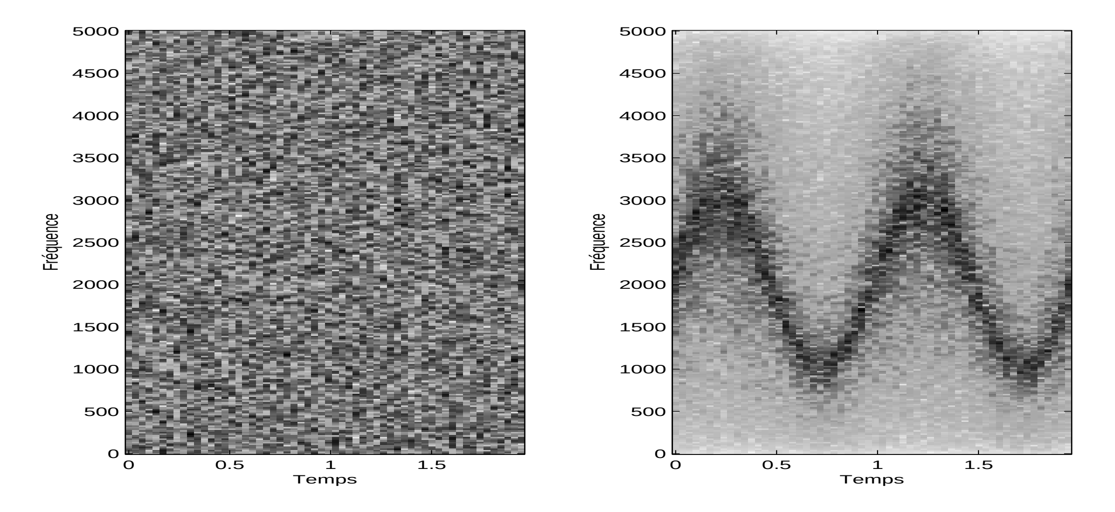


FIG. 6.6 – effet *wah-wah* sur un bruit blanc. La fréquence médiane f_m du filtre passe-bande varie entre 1000 et 3000Hz, avec une largeur de bande $f_m/10$.

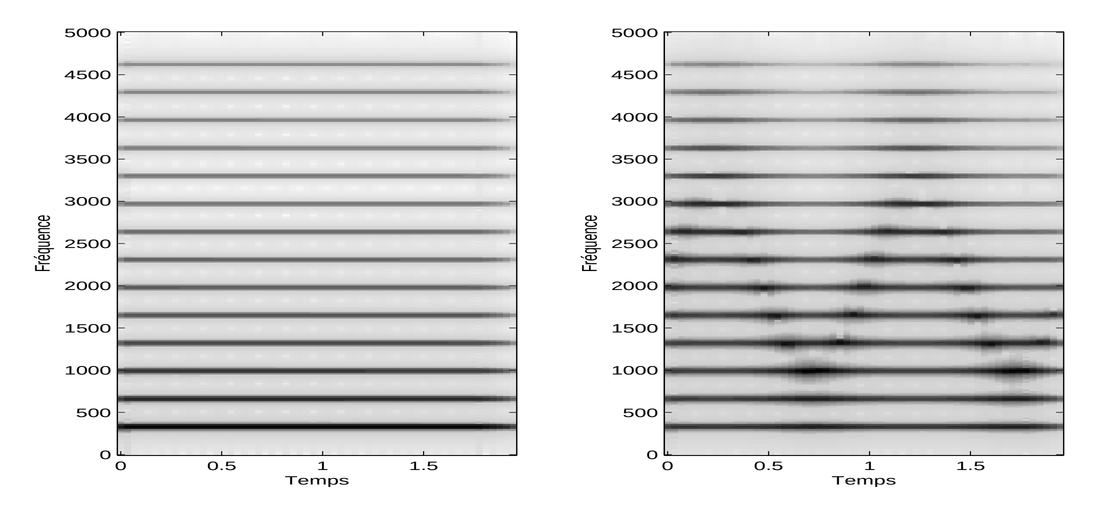


FIG. 6.7 – effet *wah-wah* sur un son périodique. Mêmes paramètres que précédemment.

6.4 Effets sonores de type AM ou FM

Ces effets sont basés sur une variation périodique de l'amplitude :

$$s(t) = (1 + \eta \sin(2\pi f_1 t)) \sin(2\pi f t), \tag{6.9}$$

ou de la fréquence :

$$s(t) = \sin(2\pi f t + \beta \sin(2\pi f_1 t)/f_1). \tag{6.10}$$

L'effet AM crée un battement de fréquence f_1, que nous avons rencontré à plusieurs reprises. L'effet FM produit un vibrato de fréquence f_1 et d'excursion fréquentielle β (cf. § 6.1.3). Dans l'effet Leslie, les deux se combinent avec un effet stéréo.

6.4.1 Vibrato

La voix et de nombreux instruments produisent un *vibrato*, dont la fréquence varie à peu près entre 4 et 12Hz. Il est facile de produire un tel vibrato par synthèse additive : il suffit d'écrire chaque harmonique sous la forme (6.10). La figure 6.8 représente le son suivant, de fondamental 440Hz, avec un vibrato de 5Hz :

$$s(t) = e(t) \sum_{n=1}^{9} \alpha_n \sin(880 n\pi t + 7\sqrt{n} \sin(10\pi t)).$$

Les autres paramètres sont $\alpha_n = 1000, 300, 0.01, 0.01, 0.01, 3, 0.5, 1, 4$ et $\beta = 7\sqrt{n}$. Une seule enveloppe a été utilisée, de forme trapézoïdale.

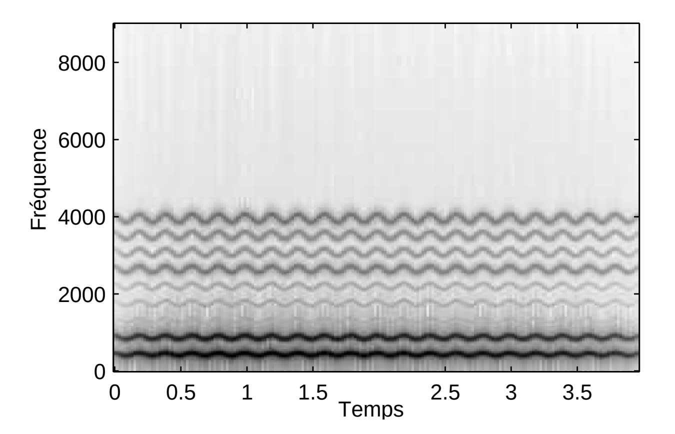


FIG. 6.8 – vibrato d'un son comportant 9 harmoniques. Cela ne vous rappelle-t-il pas quelque chose ?

En dehors de la synthèse additive, un tel effet, qui modifie la fréquence du son, peut être obtenu par une vitesse de lecture variable des données, ou son équivalent numérique. Mais il ne peut pas être réalisé au moyen de filtres, car la condition d'invariance par rapport au temps n'est pas satisfaite.

6.4.2 Effet Leslie

L'effet *Leslie* a été inventé par Donald Leslie dans les années 1940. Il fut utilisé dans les orgues électroniques Hammond, Baldwin ou Wurlizer, mais fut également appliqué aux voix, comme dans *Blue Jay Way* des Beatles. Il est réalisé par deux haut-parleurs opposés mis en rotation, produisant un effet Doppler (cf. § 6.5.1) couplé avec une variation de l'intensité.

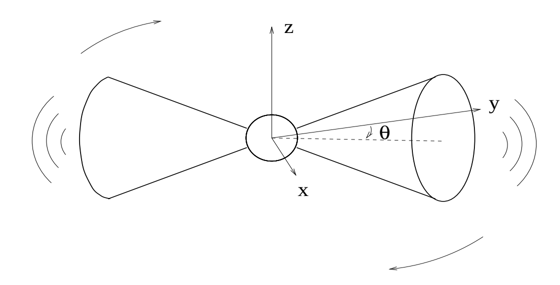


FIG. 6.9 – effet Leslie produit par deux haut-parleurs en rotation. Il fut rendu populaire par les orgues électriques.

Cet effet, nécessairement stéréo, consiste en une variation sinusoïdale de l'amplitude et de la fréquence, en opposition de phase entre les deux canaux gauche et droite, les variations d'amplitude et de fréquence étant sur chaque canal en quadrature de phase, comme indiqué sur la figure 6.10 pour un tour complet du dispositif. La vitesse de rotation est de l'ordre de 3 à 6 tours/s.

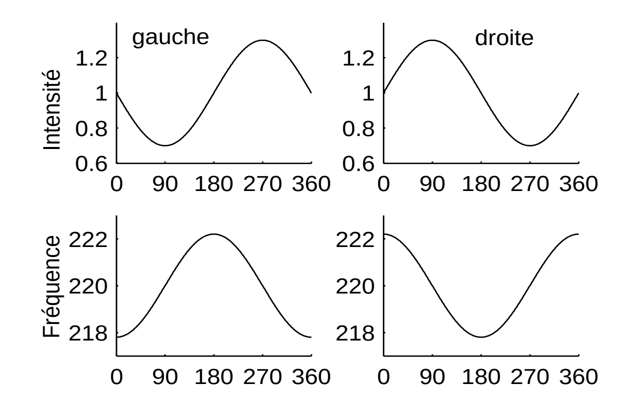


FIG. 6.10 – effet Leslie : variations d'amplitude (en haut) et de fréquence (en bas) de chaque canal, lors d'un tour complet des haut-parleurs.

On peut implanter un tel effet en procédant par synthèse additive des canaux gauche et droite $s_g(t)$ et $s_d(t)$. Pour cela, on écrit chaque harmonique sous la forme modulée en

amplitude et en fréquence :

$$s_g(t) = \sum_{n\geq 1} e_n(t)(1-\eta\sin(2\pi f_L t))\sin(nf(2\pi t - \mu\sin(2\pi f_L t)/f_L)),$$
$$s_d(t) = \sum_{n\geq 1} e_n(t)(1+\eta\sin(2\pi f_L t))\sin(nf(2\pi t + \mu\sin(2\pi f_L t)/f_L)),$$

où f est la fréquence du fondamental, $e_n(t)$ est l'enveloppe de chaque harmonique, et par exemple $\eta = 0.3$, $\mu = 0.01$ et $f_L = 4$. À noter que les fréquences instantanées sont (en faisant abstraction de l'effet produit par la variation d'amplitude)

$$f_{\text{inst}(g,n)} = nf(1-\mu\cos(2\pi f_L t)),$$
$$f_{\text{inst}(d,n)} = nf(1+\mu\cos(2\pi f_L t)).$$

Le spectrogramme de chaque canal est représenté sur la figure 6.11. Pour la visibilité, les valeurs de μ et η ont été exagérées ($\eta = 1$ et $\mu = 0.1$).

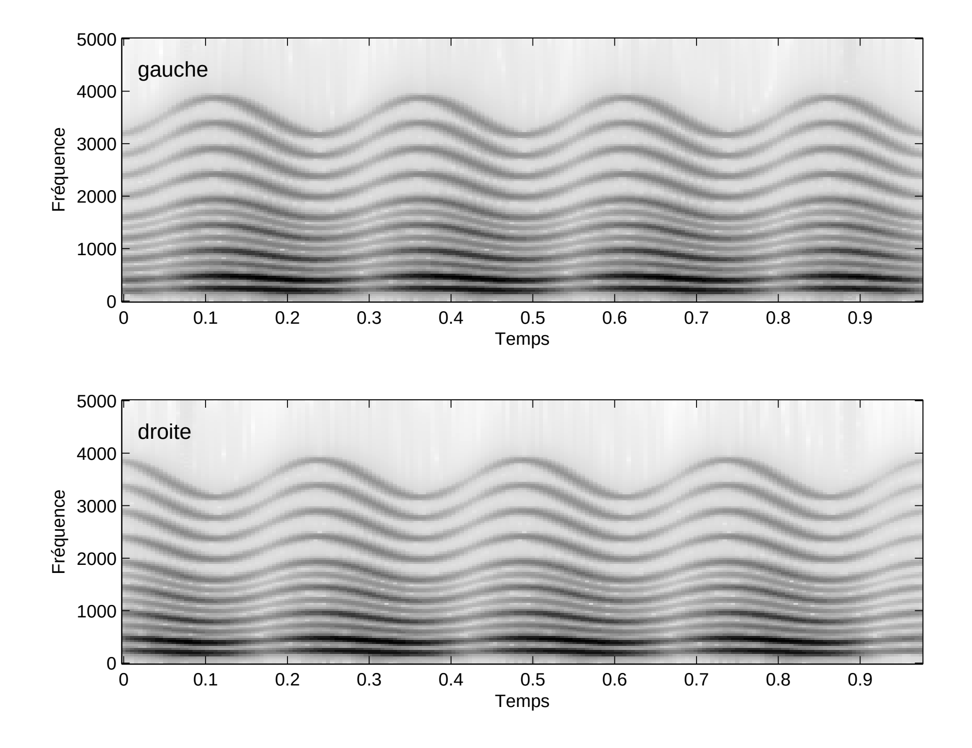


FIG. 6.11 – spectrogrammes de l'effet Leslie (on a un peu forcé le trait...)

6.5 Sujets d'étude

6.5.1 Effet Doppler (**)

Une source sonore ponctuelle harmonique de fréquence f, placée à l'origine, génère une onde de pression de la forme

$$p(x, y, z, t) = \frac{\exp(ikr - 2i\pi ft)}{r}$$

avec $r = \sqrt{x^2 + y^2 + z^2}$. Dans les deux cas suivants, calculez et représentez graphiquement la fréquence instantanée perçue par un auditeur qui se déplace :

- avec un mouvement rectiligne uniforme (sa position est donc de la forme $M(t) = M_0 + tV$, $V = (V_1, V_2, V_3)$ étant le vecteur vitesse et M_0 sa position à l'instant $t = 0$),
- en décrivant un cercle avec une vitesse angulaire constante.

Dans le premier cas, on établira les limites des fréquences instantanées pour $t \to \pm\infty$, et l'on étudiera aussi le cas particulier où les vecteurs M_0 et V sont colinéaires.

6.5.2 FM et Chowning (***)

Soit le son modulé en fréquence

$$s(t) = \sin(2\pi ft + \beta \sin(2\pi gt)/g).$$

Nous allons effectuer l'analyse fréquentielle de ce son, et déterminer les cas intéressants où il est périodique. On utilisera les deux formules trigonométriques suivantes :

$$\sin(a + b) = \sin a \cos b + \sin b \cos a,$$
$$2 \sin a \cos b = \sin(a + b) + \sin(a - b)$$

1) On pose $\mu = \beta/g$. Vérifiez que

$$s(t) = \sin(2\pi ft) \cos(\mu \sin(2\pi gt)) + \cos(2\pi ft) \sin(\mu \sin(2\pi gt)).$$

On remarque que

$$\exp(i\mu \sin(2\pi gt)) = \cos(\mu \sin(2\pi gt)) + i \sin(\mu \sin(2\pi gt)).$$

La fonction de Bessel de première espèce (rencontrée à propos des membranes) peut s'écrire

$$J_n(\mu) = \frac{1}{2\pi} \int_0^{2\pi} \exp(i(\mu \sin x - nx))\, dx. \tag{6.11}$$

Montrez que le développement en série de Fourier de $\exp(i\mu \sin(2\pi gt))$ est le suivant :

$$\exp(i\mu \sin(2\pi gt)) = \sum_{n=-\infty}^{+\infty} c_n \exp(2i\pi ngt)$$

avec

$$c_n = J_n(\mu).$$

2) Vérifiez que $J_n(\mu)$ est réel, puis, en effectuant le changement de variable $y = \pi - x$ dans (6.11), vérifiez que $J_n(\mu) = (-1)^n J_{-n}(\mu)$. Déduisez-en que

$$\cos(\mu \sin(2\pi gt)) = c_0 + 2\sum_{n\geq 1} c_{2n}\cos(4\pi ngt),$$
$$\sin(\mu \sin(2\pi gt)) = 2\sum_{n\geq 0} c_{2n+1}\sin(2\pi(2n+1)gt),$$

et montrez finalement que

$$s(t) = J_0(\mu)\sin(2\pi ft) + \sum_{n\geq 1} J_n(\mu)\left[\sin(2\pi(f+ng)t) + (-1)^n \sin(2\pi(f-ng)t)\right].$$

3) Nous faisons ici abstraction du fait que certains $J_n(\mu)$ peuvent être nuls. Que peut-on dire de $s(t)$ – ce son est-il périodique ? Si oui, quel est son fondamental ? Quels sont les harmoniques présents ? – dans les cas suivants :

- $g = f$;
- $g = f/q$ avec $q > 0$ entier ;
- $g = pf/q$ avec $p, q > 0$ entiers et premiers entre eux (réponse : le fondamental est f/q).

6.6 Travaux pratiques sur ordinateur

6.6.1 Synthèse de sons

Enregistrez et analysez (temps, fréquence et temps-fréquence) une note d'un instrument de votre choix. A partir de cette analyse, créez, sur le modèle de la fonction `instrument` écrite lors des TP du chapitre 2, une fonction qui imite le mieux possible le son que vous avez enregistré.

6.6.2 Synthèse de Chowning

Toujours sur le modèle de la fonction `instrument`, créez une fonction nommée `chowning` ayant pour en-tête

```
function s = chowning(f0,T,Fe)
```

et délivrant un son de la forme

$$s(t) = \sin(2\pi f_0 t + \beta \sin(2\pi f_1 t)/f_1).$$

Les valeurs de f_1 et β seront spécifiées à l'intérieur même de la fonction. Nous conservons ainsi le format général des fonctions instruments, ce qui permet de jouer `chowning` *via* la fonction `play` sans modifier celle-ci. Testez différentes valeurs.

Pour créer un son stéréo, vous pouvez créer un canal gauche `sg` et un canal droit `sd` selon le modèle précédent, puis les réunir en une seule matrice reconnue par la fonction `sound`. Les

commandes sont les suivantes :
calcul de `sg`
`f0 = f0+0.8 ;`
calcul de `sd`
`s = [sg ;sd]' ;`
A noter la légère modification de fréquence avant de passer au calcul de `sd`, en vue de produire un effet de phase. Explication de la dernière ligne : `sg` et `sd` sont deux vecteurs lignes (*a priori*, c'est ainsi qu'on les a programmés), et le `;` qui les sépare indique qu'ils sont rangés l'un sous l'autre dans une matrice à deux lignes et autant de colonnes que `sg` a d'éléments. Cela nécessite que `sg` et `sd` aient le même nombre d'éléments. Enfin, le 'prime' transpose cette matrice pour qu'elle soit sous la forme reconnue par la fonction `sound` dans le cas de sons stéréos.

6.6.3 Réverbération

Le réverbérateur de Moorer est programmé dans le fichier `reverb.m` qui se trouve sur ma page web. Télé-chargez le, lisez-le, et faites le lien avec la description faite en cours. Cela vous permettra de jouer sur les différents paramètres (retards, temps de réverbération, etc.). Utilisez cette fonction à l'intérieur de l'une vos fonctions instruments pour en modifier la sonorité.

6.6.4 Vibrato

Utilisez la formule décrite dans le cours pour ajouter du vibrato à votre instrument. Pour cela, le plus pratique est de modifier la fonction `synthad.m` : enregistrez la sous un nouveau nom, `synthadv.m` par exemple, et faites les modifications dans ce nouveau fichier en vous appuyant sur les formules du cours. Pour plus de réalisme, vous pouvez faire démarrer le vibrato une demie seconde après le début du son.

6.6.5 Effet Leslie

Ouvrez un nouveau fichier sur le modèle `instrument.m`, que vous appellerez `leslie.m`, et dans lequel vous programmerez un son d'orgue (rappelez-vous qu'une note d'orgue est typiquement composée de plusieurs tuyaux, dont les fondamentaux sont en progression harmonique, typiquement 1, 2, 3, 4, 5). A l'intérieur de cette fonction, vous ferez appel à une nouvelle fonction à créer `synthadl.m` (sur le modèle de `synthad.m`), qui fournira en sortie un son stéréo (cf. § synthèse de Chowning pour le format d'un tel son), et dans laquelle chaque « harmonique» sera programmé sur le modèle décrit dans le cours. L'enveloppe sera programmée dans la fonction `synthadl.m` elle-même. Utilisez ensuite le programme `play` pour jouer de votre nouvel instrument.

Index

Bibliographie

[1] S. Cordier. Piano bien tempéré et justesse orchestrale, Buchet / Castel, 1982.

[2] P. Bailhache. Une histoire de l'acoustique musicale, CNRS Editions, 2001.

[3] L. Fichet. Les théories scientifiques de la musique, Librairie J. Vrin, 1996.

[4] N.H. Fletcher et T.D. Rossing. The physics of musical instruments, Springer-Verlag, 1991.

[5] S. Jargy. La musique arabe, collection *Que sais-je*, PUF, 1971.

[6] J. Jouhaneau. Notions élémentaires d'acoustique, Électroacoustique, Éditions TEC & DOC, 2000.

[7] H. Junghanns. Der Piano und Flügelbau, Verlag Das Musikinstrument Frankfurt, 1979.

[8] W.M. Hartmann. Signals, Sound, and Sensation, Springer-Verlag, 1998.

[9] P. Lascaux et R. Théodor. Analyse Numérique matricielle appliquée à l'art de l'ingénieur, Masson, Paris, Milan, Barcelone, 1994.

[10] E. Leipp. Acoustique et musique, Masson, 1980.

[11] M.V. Mathiews. The Technology of Computer Music, M.I.T. Press, 1969.

[12] N. Moreau. Techniques de compression des signaux, Masson, 1995.

[13] J. Pierce. Le son musical, bibliothèque Pour La Science, Belin, 1999.

[14] J.C. Risset. Hauteur, harmonie, timbre, synthèse, in Musique, rationalité, langage - l'harmonie : du monde au matériau, L'Harmattan, Paris-Montréal, 153-167, 1999.

[15] E. Zwicker et R. Feldtkeller. Psychoacoustique, l'oreille récepteur d'information, Masson, 1981.

[16] Colloque acoustique et instruments anciens, factures, musiques et science. Musée de la musique, 1998.

[17] DAFX - Digital Audio Effects, John Wiley Sons, 2002

[18] Proceedings of the International Symposium on Musical Acoustics, July 2-6, 1995, Le Normont, Dourdan, France, Ed. SFA.

[19] Encyclopédie Universalis

[20] Les instruments de l'orchestre, bibliothèque Pour La Science, Belin, 1995.

[21] Sons et musique, bibliothèque Pour La Science, Belin, 1979.